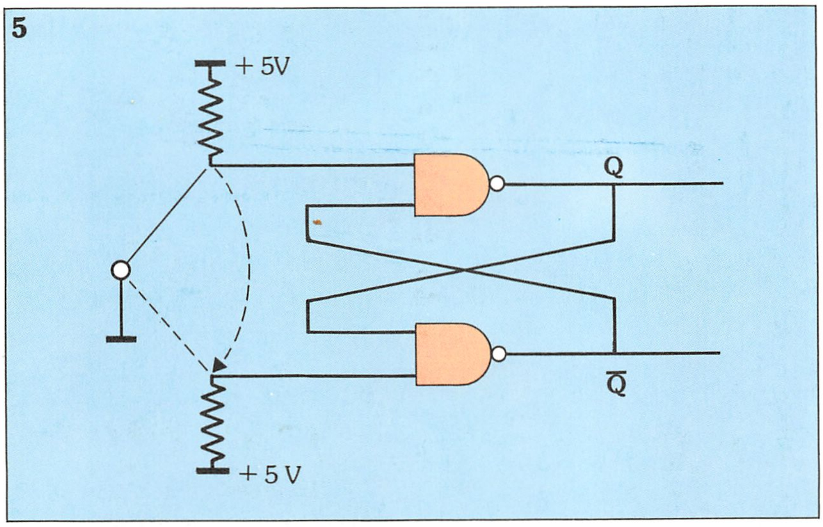
(continued from Part 6)

Integral memory

As explained previously, the advent of semiconductor techniques (particularly LSI) meant a rapid decline in the use of ferrite core memory. The semiconductor memories which replaced ferrite core are divided into two types: Random Access Memory (**RAM**) and Read Only Memory (**ROM**).

As the name implies, RAM can be accessed directly i.e. without the necessity of sifting through other data. Ferrite core memory is one example of this, but the term RAM is now used to describe any memory from which it is possible to read and write. For example, where a home computer is advertised as having 16K RAM it means that 16 kbytes can be available for the user to load and access data of his choice.

ROMs are able to carry out reading programs only.



5. A simple flip-flop circuit, which can act as a 1-bit memory.

RAM semiconductors

A RAM semiconductor is a volatile read/write memory, differing from a permanent core memory in that data is lost as soon as the power is switched off. Today's LSI techniques mean that large numbers of memory cells, along with all the necessary selection, reading and writing circuits, can be built on one chip, with excellent access times.

There are 64K RAMs available today which measure between 14 mm^2 and 17 mm^2 offering average access times of around 50 to 80 ns.

Static and dynamic RAM

RAMs are either static or dynamic. **Static** RAM has a flip-flop as its elementary memory cell. Flip-flop circuitry is covered in detail in the *Digital Electronics* section, but essentially flip-flop circuits are bi-stable i.e. they have two stable output states corresponding to logic levels 1 and 0. The conversion from one state to another occurs only as a result of the correct input command from the input switch.

Figure 5 shows that the position of the input switch determines a stable logic level (1 or 0) at output Q. The state does not change unless the switch is moved to the opposite position. Output Q can, therefore, represent a memory bit.

This type of memory is termed static because of its ability to indefinitely retain the logic levels memorized, as long as a regular power supply is maintained.

Static should not be confused with the concept of a non-volatile state. Static memories *are* volatile in that they retain the data only when there is power.

Static memories can be costly, even in semiconductor terms, so another method of constructing semiconductor memories has been developed which gives a higher degree of integration and a significant reduction in production costs. These are called **dynamic** memories. Data is stored by these memories in the form of a charge in the input capacitance of a transistor.

Logic levels 0 and 1 are represented by the state of charge or discharge of this capacitance. However, since the insulation of this device is not perfect, the charge stored will tend to flow away in milliseconds and so cause the loss of the data. To avoid this, the charge of the capacitor must be periodically regenerated, which is done by means of an operation called **refresh.**

This consists of reading the information registered in a certain area of the memory and rewriting it in the same area. The refreshment is carried out automatically at time intervals preset on each memory cell. The refresh 'sweeps through' the

memory in that different parts are refreshed in sequence during different phases. It happens as a continuous process as long as the computer is switched on.

The need to refresh has two disadvantages: first, additional logic is needed for this operation making the circuits bulkier; second, there is an obvious increase in the time taken for all the memory operations.

However, compared with static chips, the structure of dynamic chips is much simpler. Also the length of refresh time is negligible compared with the overall processing time, for most applications.

There still remains, both for static and dynamic RAMs, the problem of volatility. This problem becomes serious when the computer is in charge of the control of industrial plant, or of any delicate application. In these cases, a temporary breakdown in the power supply, even if extremely brief, could cause the loss of the program. This would mean that the computer could not take control of the process again when the supply returned.

The simplest and most common solution is to provide batteries for temporary breakdowns (back-up batteries). These are generally bulky and expensive – the greater the length of time that they are expected to provide power, the bigger they have to be. The cost is usually justified in that it would be much more expensive to restore a disrupted industrial process. Normally however, these batteries are expected to last for a short period of time. In the case of prolonged power failure it is taken for granted that the process being controlled by the computer will, in any case, come to a halt.

ROM semiconductors

ROMs are non-volatile memories which retain data regardless of the presence of power. But, once they are programmed and inserted in the computer system, they can only be read. They are simpler and less costly than RAMs, and their big advantage is in not being volatile. But, as they cannot be modified by a program, they can only be used for particular applications.

A semiconductor ROM is, in effect, a matrix (or lattice) of conductors organized in lines and columns. Each crossing point between a line and a column represents a bit at logic level 0 or 1 depending on whether there is electrical contact between the line and the column. The ROM is programmed by closing or 'making' these contacts. With large mainframe computers, this is generally done specifically for the user when it is built. It is irreversible and in technical terms is said to be 'burned' onto the memory. A **template** is the configuration of bits which represent the program. Contacts are made according to this. Production costs are high for this type of ROM and so a certain minimum quantity of units must be made for their production to be economic.

There are a number of different types of ROM with slightly varying characteris-

Science Photo Library – Jerry Mason

Floppy disks are widely used in microcomputers as they are small, cheap and easy to change. The disk is simply slotted into the drive unit.

tics. These are described below.

Programmable read only memory (PROM)

This is a memory chip which can have a program permanently implanted into it. The programming is achieved by using a special device called a PROM programmer. It is usually used as a component of a computer to carry the permanent instructions it needs. Once the program is implanted it can never be erased.

Erasable/programmable read only memory (EPROM)

Again, a program can be implanted which will be incorruptible as long as it is in the system. The program is held in such a device by charging or not charging specific memory cells within the matrix to form the required 0 and 1 logic levels.

All information programmed into an EPROM can be erased if required by exposing the top surface of the package to ultraviolet radiation. This brings about an ionising action within the package which causes each memory cell to be discharged. The matrix can then be used again to store a new program. The new program would be implanted using a PROM programmer.

Electrically erasable programmable read only memories (EEPROM)

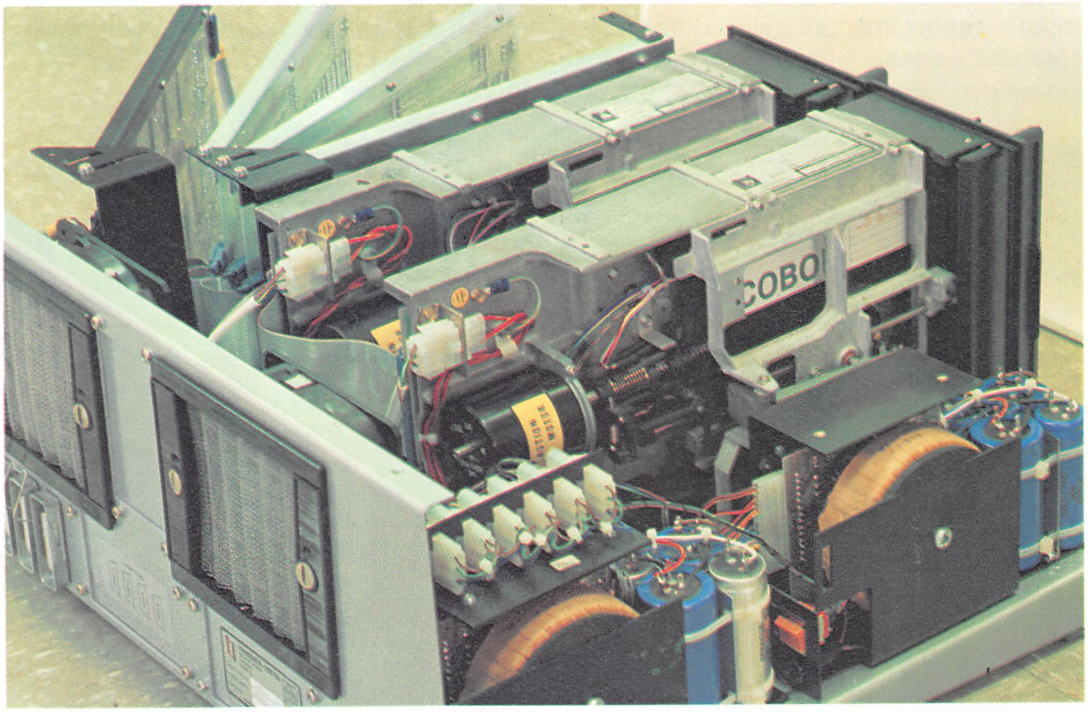
These are a very recent development, similar to EPROMs, but erasable electrically. This greatly increases the speed of erasing and reprogramming is achieved by applying a strong current pulse removing the entire program, leaving the device ready to reprogram.

Electrically alterable read only memory (EAROM)

This is usually used in military and high precision industrial applications. A more accurate description of EAROM would probably be 'mainly reading' memory rather than 'read only' as it can in fact be written on. However, the writing operation is slower than the reading process (several milliseconds as opposed to one microsecond), uses complex techniques allowing only low integration levels, and demands multiple power supplies. This means that EAROM cannot be used as normal read/write memory.

Magnetic film

This is a different type of central memory, sometimes referred to as 'thin film'. The elementary memory cell is made up of spots of magnetic material a few milli-



Reverse view of a twin floppy disk drive, showing power supply, motor drive and control circuitry.

Science Photo Library – Jerry Mason

metres in diameter and a few tenths of a micrometre thick. The magnetic film is placed on a vitreous support and subjected to a strong magnetic field which produces a preferential magnetic direction in the **dipoles** which make up the spots. The elementary dipoles are all lined up according to the 'preferred' direction, in one way or the other. One direction corresponds to logic 0 and the other to 1.

We won't go into the operation of this type of memory in detail but, like ferrite core, it can also be organized in matrices, with several planes of spots, and the memory word made up of two spots per plane.

Although large sums of money were spent on the development of various forms of magnetic film memory, they were overtaken by the rapid growth of semiconductor memories. As these are both faster and more compact, magnetic film could not compete, and this type of memory is no longer used in any computers.

Mass memory

Mass memory is the term given to all memory that is **peripheral** or auxiliary to the computer. It is used to store large quantities of data as well as libraries of programs, and is usually in the form of magnetic disk or tape. In the case of data, the computer will access it directly from the mass memory medium on command from the user. With a program the user will 'load' the program from the disk or tape into the computer's central memory.

Of course, mass memory has to be non-volatile, as it is required to store data for long periods without power supply.

The main differences between various types of mass memory device lie in the way in which the magnetic material is contained – be it on plastic tape, plastic disk, metal disk, or in a drum. We'll take each device in turn and look at their individual characteristics.

6. The basic arrangement of a magnetic tape drive unit.

7. Read/write head from a tape drive.

8. A 9-track magnetic tape, showing the layout of the data records.

A suitcase-style microcomputer. To the far right of the front panel are two Winchester disk drives used for mass storage, one fixed and one with the unusual feature of accepting a removable Winchester cartridge.

Jonas International

Magnetic tape units

Home computer users will be familiar with the use of magnetic tape as a data and program storage medium. Ordinary audio-cassette recorders can be used in conjunction with most small machines. They work in exactly the same way as sound recordings, except that the signals stored on the tape represent binary numerals. So when you play the tape, you can't hear anything intelligible, but the computer can 'read' it.

Tapes have, in fact, been used for storing computer data for years. The tape units used with large, commercial mainframe computers are complex electronic and mechanical systems (see *figure 6*). They have elements which start and stop the two reels on which the magnetic tape is wound: these must respond at very high speed because to search for and find data the program will frequently require the tape to stop, start and reverse. Also, when the unit is reading or writing data the tape has to be kept at a constant tension. This is achieved by means of partial vacuum columns under the reels.

Figure 7 depicts a read/write head. (Some equipment has a read head separate from the write head.) A **head** is a device which causes magnetization of tiny sections of the tape during writing, and receives induction signals during reading. In principle it works in the same way as the head in a domestic tape recorder.

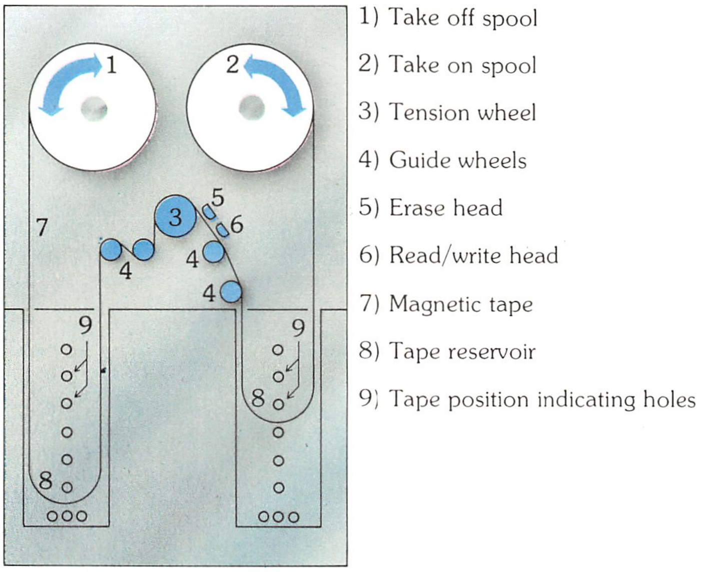
A section of the tape is magnetized in one direction to register logic 0 and in the other to register 1. The current induced in the head in the reading phase is different in each case, allowing 0 and 1 to be read. This type of reading is not destructive (i.e. the data is not lost from the tape).

The surface of a tape is magnetized along longitudinal channels called **tracks.** *Figure 8* shows a nine-track tape, being one byte (one bit per track), plus one bit (on the ninth track) for the **parity control.**

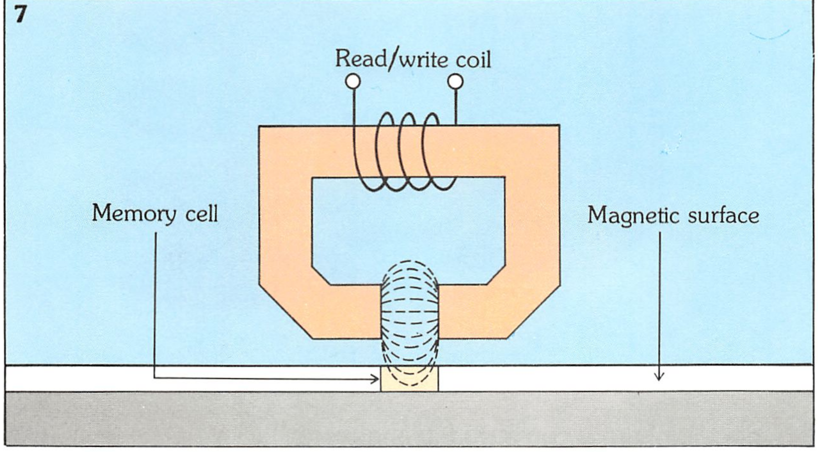
The **recording density** on computer tape can vary, the most common being 256, 512, 800 and 1600 characters per inch. Memory capacity for magnetic tape depends on its length, which is usually 350, 700 or 1050 metres. The actual quantity of data which can be recorded depends on a further factor, which we will look at now.

Normally, a magnetic tape is stationary. The program generates a signal for it

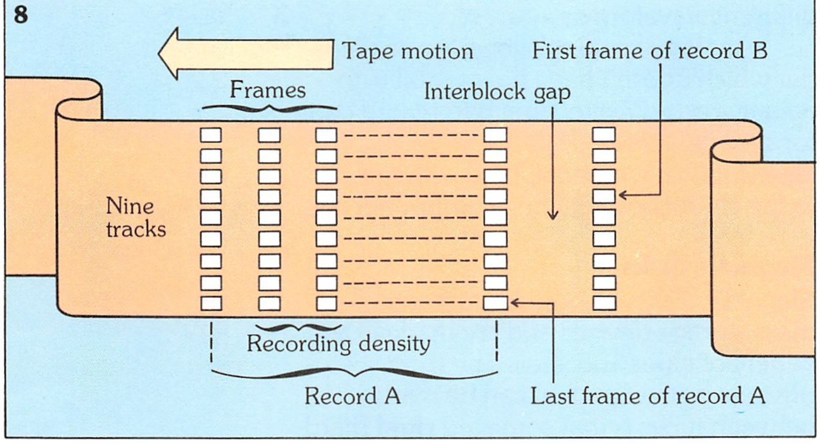
6



7



8



to start when something has to be recorded. Once the running speed is constant, then recording can begin. The piece of tape which has moved in the meantime does not contain recordings, and is known as the **interblock gap** or **interblock space.**

The recording takes place under the control of the program for a given number of characters (the maximum is determined by the size of the computer's central memory), after which another gap is created. This means that the recordings of data on the tape are divided into blocks of characters, and between each block there is a gap, formed partly by the end of one piece of recording and the beginning of the next. If each block contains only a few characters then there is a wastage of tape.

Every read or write operation on the tape involves a data block which must fit into the computer's central memory as that is all the computer can deal with at a given moment.

Sequential access

Magnetic tape allows only **sequential access,** which means that to find a piece of data all other previously recorded data must be run through. Moreover, a magnetic tape can never be used for reading and writing simultaneously. If the user wishes to change the content of a tape, it must be read in blocks into memory (i.e. of the computer). The change can then be made – a block at a time – and re-copied onto a new tape.

The smallest working unit of a magnetic tape is a block, so any read or write operation involves the transfer of an entire block of data from the tape to the central memory and vice versa. In this case, **access time** (the time taken by the computer to read the data) cannot be predicted, as it depends entirely on the position of the block on the tape.

The total recording capacity of a tape can be very large. For example, a tape 700 metres long with a recording density of 340 characters per centimetre can contain about 24 million characters.

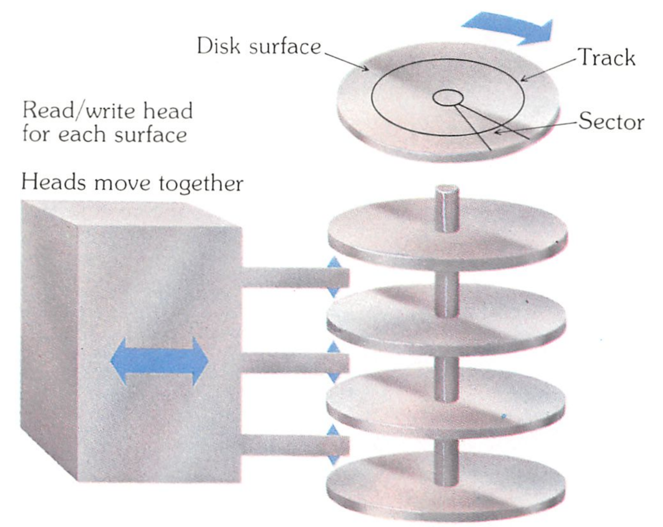
Inhibiting writing

For reasons of security, it is not always desirable for data to be written onto tape, or for existing data to be changed, by anybody who happens to be using the computer. It is possible to prevent or **inhibit** writing by physically placing a protective ring on the reel or by using a special switch on the tape unit. The tape has a reflecting strip at the beginning and end to mark which part of it can be used to store data.

As we mentioned earlier, tapes of a much less complex nature are used in mini and microcomputers. Unlike those used for large scale computers, these tapes are **single-track cassettes** and are less reliable by comparison. Data can be either recorded digitally onto cassette in the way

9. The basic arrangement of a fixed disk drive, showing the recording heads.

9



described earlier, or converted to analogue signals with logic 0 and 1 having two different waveforms.

Analogue coding allows a tape recorder which doesn't have the capability to reliably record *digital* signals to record data for storage. Unlike disks the recording on tape is, of course, serial, both with regard to the characters and their component bits.

Magnetic disks

Magnetic disks are popular non-volatile mass storage devices and are made in a variety of types and sizes. The most obvious distinction that can be made between types is that some are **rigid** (hard

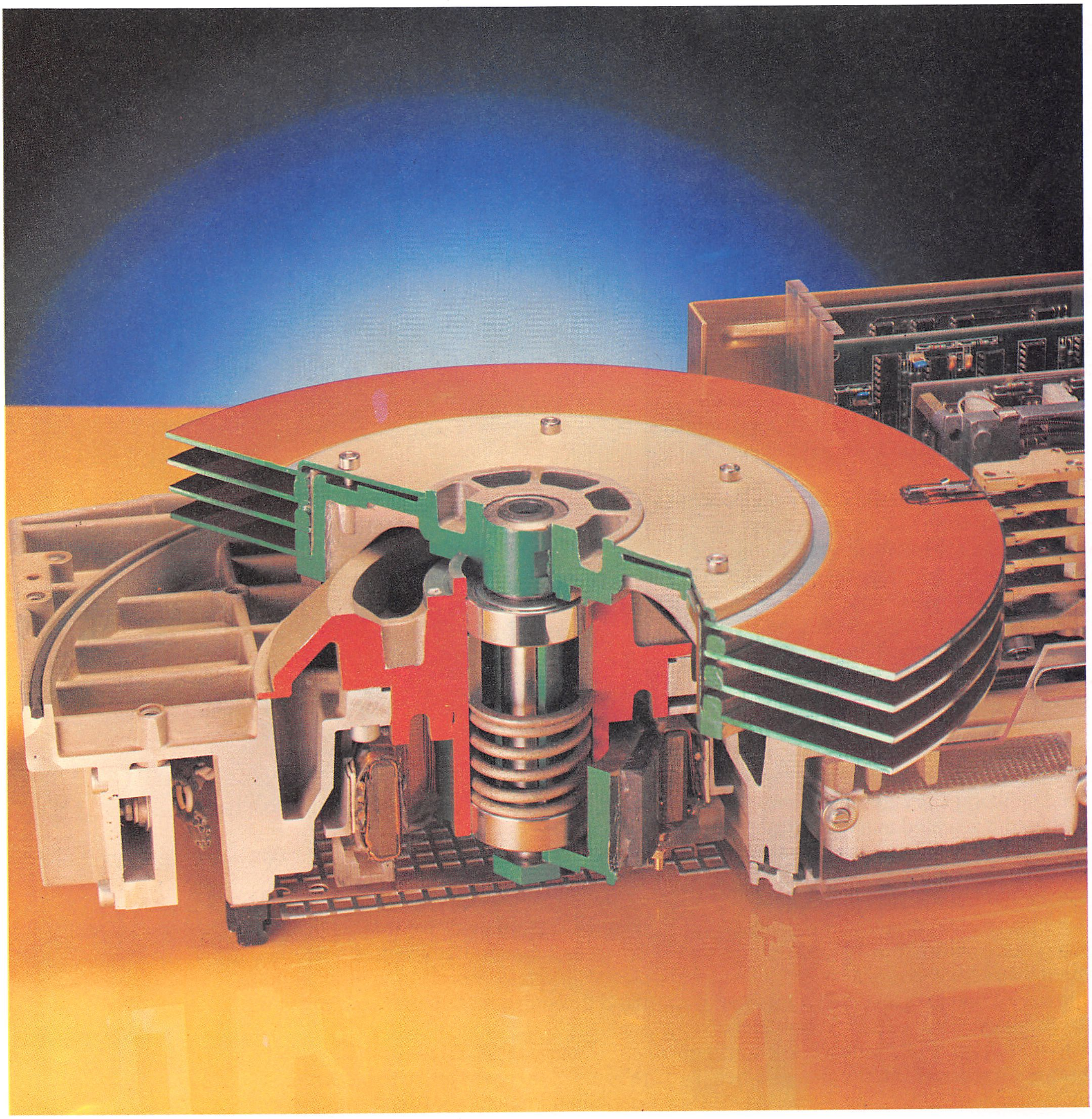
A cutaway view of a **Winchester disk drive,** showing the motor heads, and control circuitry.

disks) and some **flexible** (floppy disks). Both types basically consist of a rotatable disk that has a coating of magnetic material (like that on recording tape) on one or both of its sides. The main advantage of disk data storage over that of magnetic tape is that the time taken to find and read a piece of data (**access time**) is considerably faster. Disks can therefore be considered as **direct access** memories.

Rigid disks

Rigid disks can either be fixed – that is to say permanently mounted in their drive units – or they can be removable, meaning that disks and the information recorded on them can be changed whenever necessary.

In large mainframe computer installations the most common disk storage employs **exchangeable** (removable) disk cartridges. These cartridges contain six 14



The Research House/Tony Latham

inch disks mounted one above the other on a central spindle. When a cartridge is placed in the drive unit it is rotated at a speed of about 2400 RPM. Data is recorded and read from the disks by ten read/write heads which are mounted on five movable arms. These move in and out and enable the heads to cover all the disks' usable area. The heads do not actually touch the surface of the disk, but hover on a film of air (about 3 μm thick) which is caused by the high speed of the disk's rotation.

This basic arrangement is illustrated in *figure 9*. Because all the heads move together, only one head can be used for reading or writing at any one time.

You will have noticed that the top and bottom surfaces of the disk pack are not used for data storage. This is because these are the surfaces most likely to become contaminated with dust or finger marks which could cause **drop-outs** (misrecorded or misread information). The surfaces that are used, can each store between 1.5 to 15 Mbytes of information. (1 Mbyte = 1,000,000 bytes). Exchangeable disks are nowadays used in most large scale, general purpose data storage applications.

Fixed disks, on the other hand, are usually employed when fast access to a relatively small, stable amount of information is needed. In many cases, fixed disk units have one read/write head for each track of information they carry, giving the fastest possible access time with this storage medium. The **Winchester disk** however, is becoming the most common fixed disk storage device. This unit usually has stacked disks, which can be of 14, 8 or 5¼ inches diameter. All surfaces are used for data storage and retrieval, and the whole unit is sealed in a dustproof package. The read/write heads operate in a similar way to those of the 14 inch exchangeable disk unit.

Winchester disk units are used in mainframe computers, minicomputers and even some microcomputers according to the size of the disks involved. One side of a 14 inch Winchester disk can hold between 9 to 80 Mbytes of information – a considerable increase over the standard 14 inch disk.

Floppy disks

Floppy disks are exactly what their name implies – flexible. Floppy disks are available in 8, 5¼, and 3½ inch diameter sizes, and can be either single or double sided. A floppy disk consists of a piece of plastic film about 0.5 mm thick, coated with a magnetizable material. This is fixed within an envelope (or sleeve) protecting the disk from damage. The disk is rotated by a drive shaft that fits in the centre of the sleeve, and is exposed to the read/write head through a rectangular aperture in the sleeve.

Floppy disks are reasonably cheap (about £2.00 each) and can be easily loaded and unloaded. Along with these

10. Information flow and control in a floppy disk system.

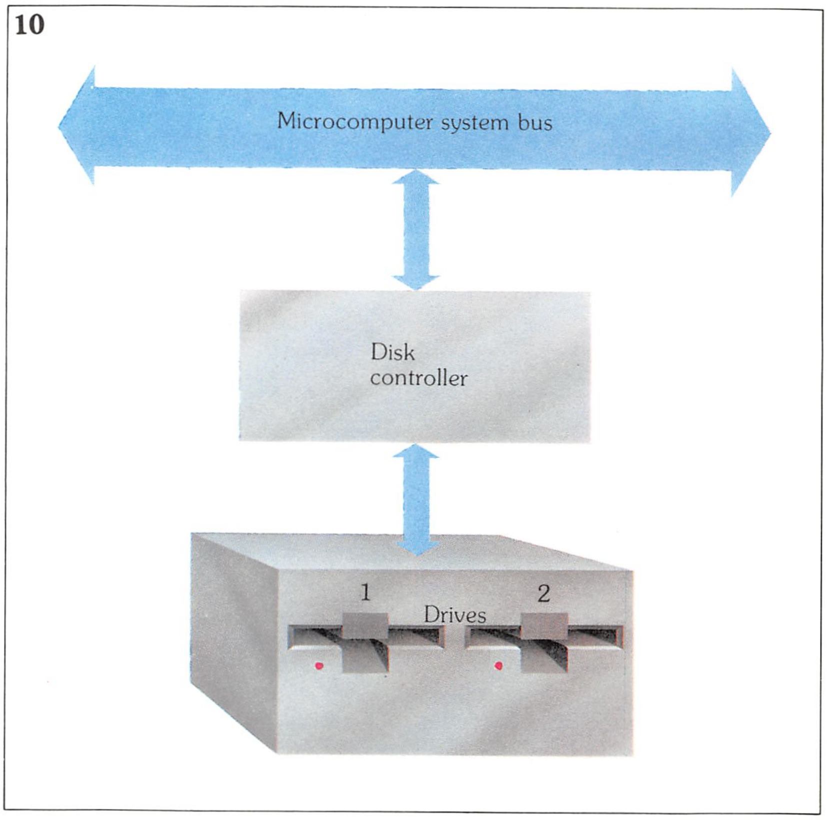


Table 1

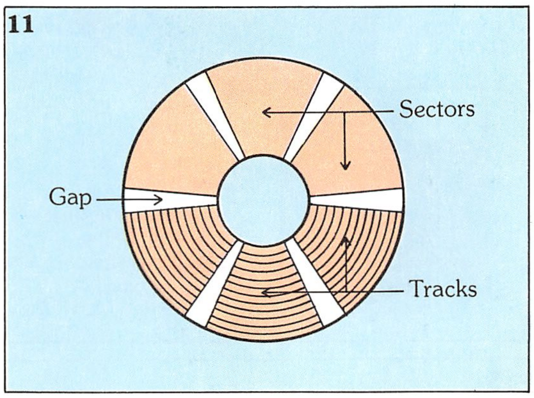
Characteristics of an 8″, single sided floppy disk

Capacity	per disk per track	256. 256 byte (formatted) 3. 328 byte (formatted)
Data transfer speed		250 kilobit/second
Access time	latency from one track to the next (10 ms to reach the sector)	10 ms
Rotation speed		360 rpm

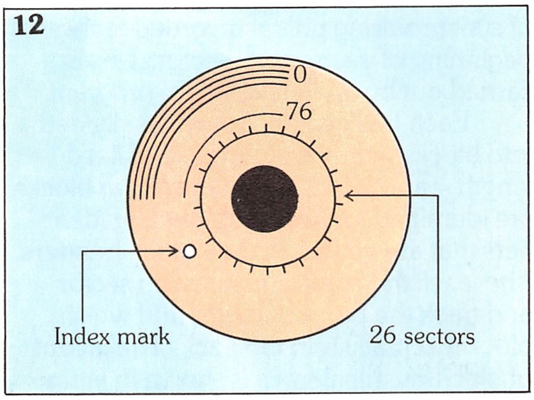
reasons, the fact that they have faster access times than cassette tapes explains the popularity of floppy disks in most of today's microcomputers. In terms of storage density, one side of an 8 inch disk can hold from 0.4 to 0.8 Mbytes of information. Like conventional disk units, floppy disk drives need controlling and driving circuitry, but because of its low cost, this hardware is more delicate and less reliable than that of a mainframe disk drive. This leads us to one crucial difference from hard disk devices – floppy disks do not rotate continuously.

Figure 10 shows how the floppy disk system operates. Two parts can be distinguished, the disk drive and the disk controller. The controller contains a microprocessor. Whether it is in a floppy disk drive or a rigid disk unit, the controller acts as an interface between the computer and the disk drive, controlling the transfer of data from one to the other. *Table 1* lists some of the characteristics of a typical 8 inch single sided, (single density) floppy disk.

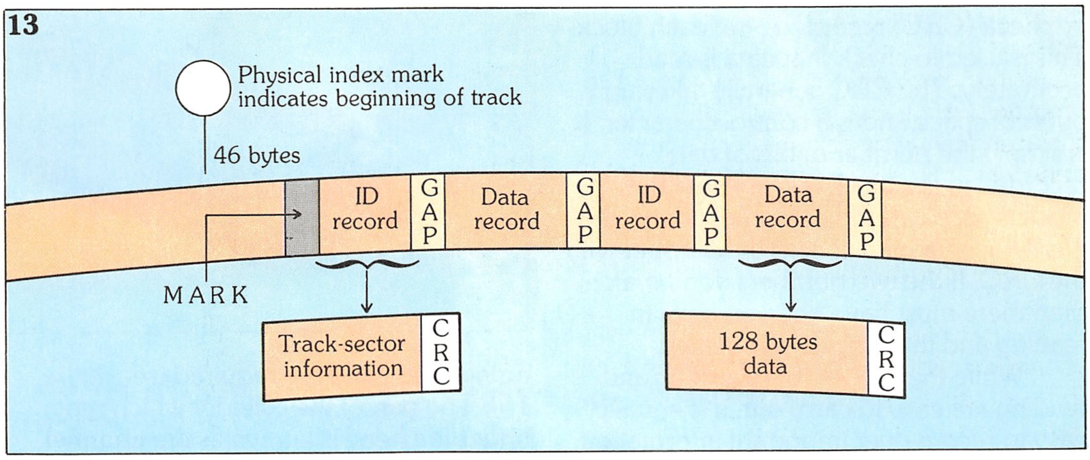
11. Disks are divided into track and sectors to enable addressable storage.



12. This disk has 76 tracks and 26 sectors. The index mark is to enable the control circuitry to know the position of the sectors at any one time.



13. Soft sectored data is recorded with an identification record before each data record. This ID record contains the address information.



Formatting and accessing

Before any information can be stored on any type of disk, the disk has to be divided into tracks and sectors, which are then formatted. This ensures that once an item of data is stored on a disk, it can also be found and read.

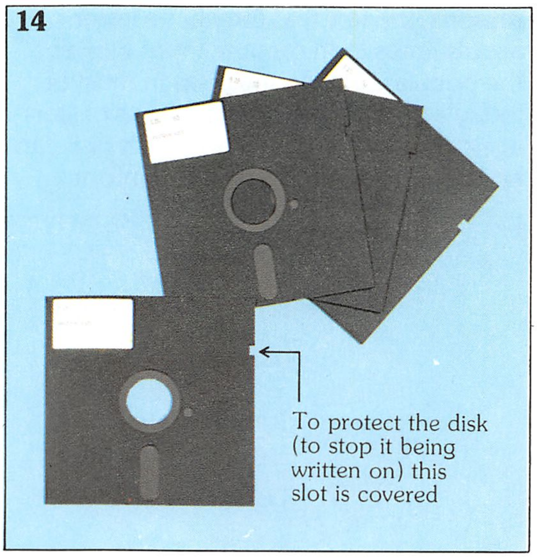
Each side of every disk is divided up by small gaps, into a number of equal sized portions known as **sectors** (*figure 11*). There can be 8, 12, 16, 24 or 32 sectors per surface, depending on the size of the disk. The disk is also divided into tracks, which of course are cut up by the sectors. A floppy disk could typically have 48 tracks per inch (measured radially) while a rigid disk could have from 100 to 400 tracks per inch (TPI). The 'standard' for exchangeable disk systems is 200 TPI, while disks that have a fixed head for every track have from 25 to 100 TPI.

So, how is data found and retrieved from a disk? Most types of disk or disk pack have an **index**, which is a mark or a hole that can be physically detected by the drive unit. The position of each sector

is thus known in relation to this index at all times. This is known as **hard sectoring** – that's to say the sectors are determined by the hardware. *Figure 12* shows the layout of an IBM 8 inch floppy disk with 76 tracks and 26 sectors. **Soft sectoring**, on the other hand, defines the sectors by a series of synchronising pulses recorded at the beginning of each actual sector. This is carried out by an 'initialization' program.

Each track within a sector is divided into blocks which are normally of fixed length – about 512 bytes. Particular blocks are identified for access by reading identifiers that are written in their block **headers**. These tell the computer in which sector and track the block belongs, and which block it actually is in the track. The layout of blocks within a track is shown in *figure 13*. This example has a cyclical redundancy check (**CRC**) recorded after each block. This is used to check that data is read accurately. The CRC generally takes up 2 bytes of space and is a control character. It is simply the number of bits of data recorded in the block – every time the block is accessed the computer counts the bits it has read and checks the number with the CRC. If the two numbers don't match then there must have been an error in reading and the process is repeated.

While the process of recording and reading are easy to carry out, it is equally easy to record over important information. To prevent this, 'inhibit' switches can be turned on in a particular drive unit, or a 'tag' can be removed or stuck onto a disk package. See *figure 14*.

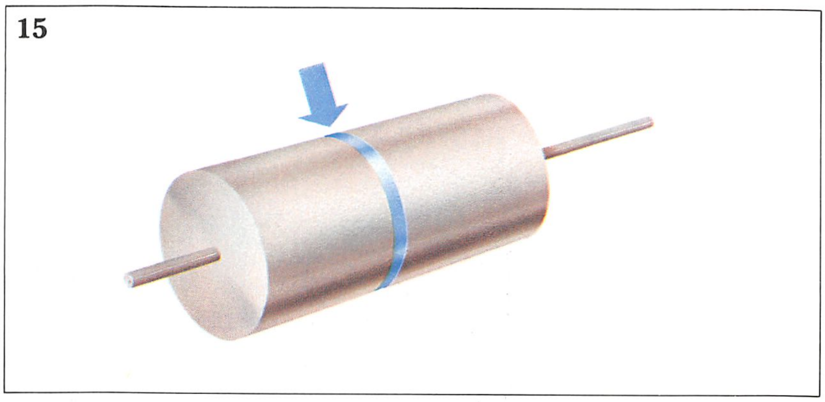


14. Recording on floppy disks can be inhibited by covering a slot on the disk's package.

Magnetic drum memories

The magnetic drum type of memory is so expensive that it is usually only used in large computers. These are direct access memories, which means that the data is recorded at both track and sector level and can be dealt with by a similar addressing system to that already described for disk units.

The drum memory consists of a metal cylinder which revolves at high speed. The surface of the cylinder is covered with a magnetic substance. The memory is accessed by a read/write head placed a few hundredths of a millimetre above the surface, meaning that very close manufacturing tolerances are required (see *figure 15)*. The part of the cylinder which runs under the head is known as the **channel** and corresponds to the track on a magnetic tape. Several channels make up a recording track on a drum (not to be confused with the track of a disk).



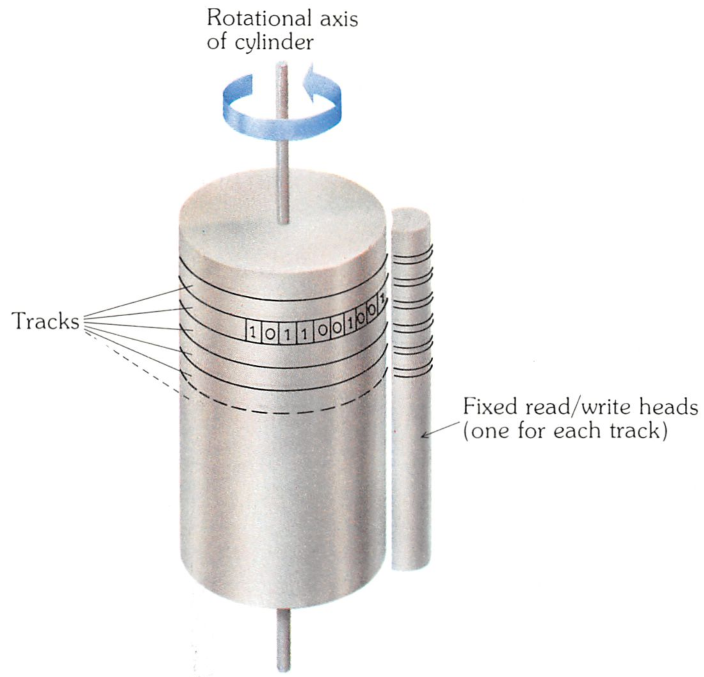
15. In a magnetic drum memory system, the drum is divided up into tracks which contain the data.

The drum behaves in a similar fashion to magnetic tape, except that the tracks have a fixed length. There is at least one read/write head for each track which can deal with all the channels simultaneously. So the bits are recorded in parallel and the characters in series. Usually there are as many heads as there are tracks. Access time depends on the location of the data within the track, but can be cut down by having several heads, one for each track *(see figure 16)*. As with disks, the recording densities of different drums vary.

Magnetic bubble memories

A recent development in the use of magne-

16

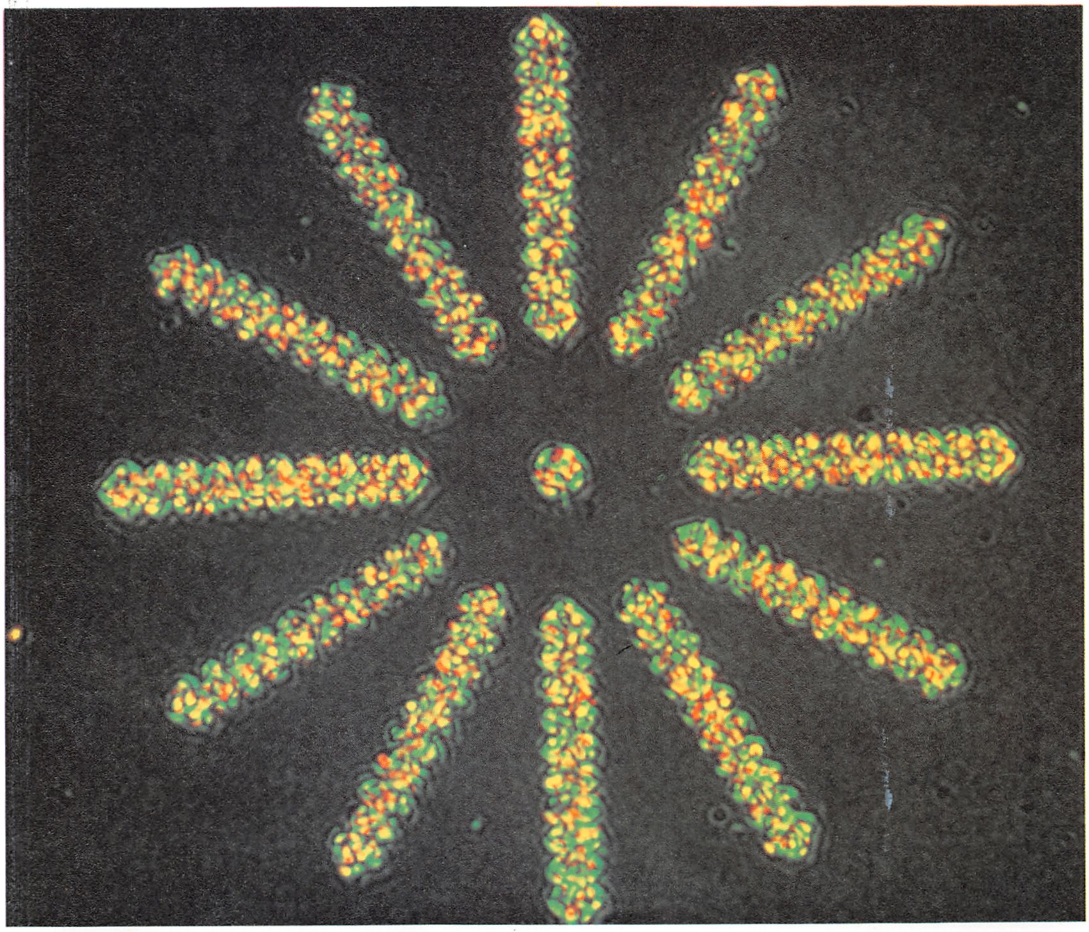


tic effects in memories is **magnetic bubble memory**. In a piece of a single magnetic oxide crystal, small areas – **domains** – are magnetized. The direction of magnetization is randomly oriented. If a magnetic field is applied, these domains may be broken up into small cylindrical magnets, arranged in a regular pattern. These tiny magnets (called **bubbles**) can move freely in the material.

Data in the form of a string of logic bits may be stored in a string of bubbles by having some bubbles magnetized upwards to represent 1 and others downwards to represent 0. This string of bubbles may be moved along a track through the material, giving a situation comparable to data stored on a magnetic tape – although here the 'tape' is stationary and the data moves within it. Since the crystal is, in effect, two dimensional, many bubble tracks may be laid out in parallel and large quantities of data can be stored. Magnetic bubble memory in some cases is being used as a replacement for floppy disks.

16. Each track of data on a magnetic drum has its own fixed read/ write head.

Magnetic bubble memory is a relatively new form of mass storage, and is still being developed to give faster access times. It is however a very compact form of mass memory. This photomicrograph shows magnetic bubbles that exist in channels one micrometre thick, which are contained in an IC. (Photo: courtesy IBM).



Glossary

access time	time taken to retrieve data from a storage medium
capacity	quantity of data which can be contained in a single memory, e.g. 64K
direct access	memory in which the data can be accessed directly, i.e. without sifting through other recorded data
dynamic memory	a memory in which data has to be refreshed at fixed intervals
formatting	setting up the magnetic material on a disk to determine the arrangement of data
interblock gap or space	the distance between blocks of data on magnetic tape
mass memory	non-volatile storage medium for data or programs, e.g. tapes, disks, etc.
memory	the data store of a computer, held on either internal or external media such as disks or tapes
parallel transfer	method of data transfer in which all the bits are transferred simultaneously
parity check	a check to see if the number of logic 1 bits in a byte or word is odd or even
RAM (random access memory)	memory in which data can be accessed at any point. Now, more usually used to mean read/write memory
ROM (read only memory)	memory containing data which cannot be corrupted, i.e. memory which can be read from but not written to
sector	part of a track of a magnetic disk or tape
sequential access	access where other data has to be sifted through before that which is required is found
serial transfer	method of data transfer in which each bit is transferred in sequence
static memory	memory which can retain indefinitely the logic levels memorised – as long as power is supplied
track	a channel for recording data on a magnetic medium: longitudinal on a tape; concentric on a disk
volatile memory	memory in which the data is lost when power is cut off

ELECTRICAL TECHNOLOGY

Capacitors

Capacitors are important components in electrical circuits. They have the ability to hold electrical charge – a phenomenon known as **capacitance.**

Figure 1 shows a battery connected up to two metal plates which are positioned facing each other a small distance apart. We know that electrons will try to flow from the negative terminal of the battery to the right hand plate, and then from the left hand plate to the positive terminal of the battery. However, as the plates are not touching, there is no electrical circuit and, after a burst of activity when the battery is first connected up, the electrons move to the plates but nothing further can happen.

However, we have produced on the metal plates the conditions for generating an electric field between them. In this field are atoms of the gases which make up air. Each atom, because it contains protons and electrons, finds itself being 'stretched', because its protons are attracted to the negatively charged plate and its electrons are attracted to the positively charged plate.

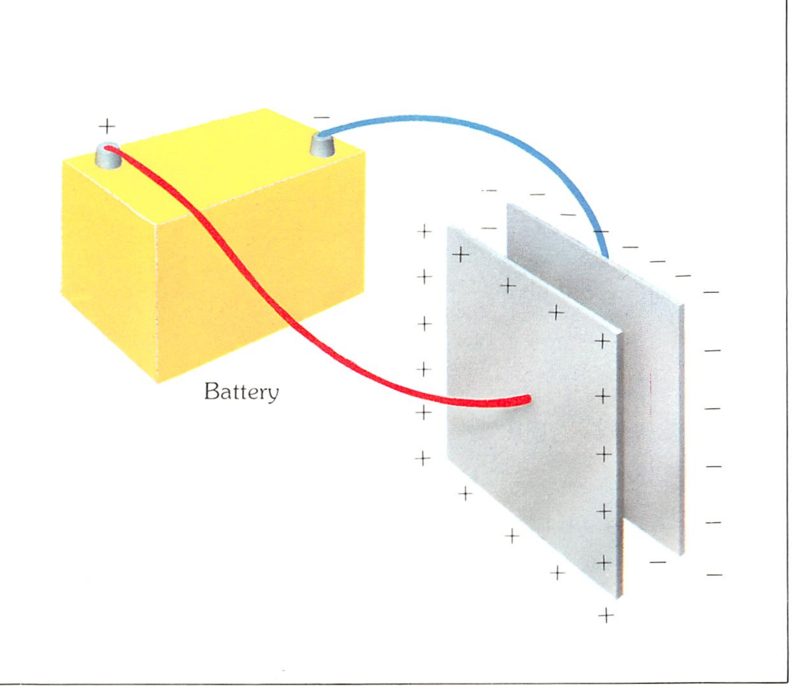
Figure 2 shows this effect. Connecting the battery effectively does work on the air. That work is now stored in the 'stretched' atoms in the form of electric charge. The plates can therefore be said to have **capacitance** – the ability to store charge. The higher the voltage from the battery, the stronger the field between the plates and the more energy is stored in the atoms.

If the battery is disconnected from the plates there are still extra electrons left on one plate and insufficient on the other, with 'stretched' (technical term **polarized**) atoms between them. The plates now form a charged capacitor. It can be discharged by touching together the wires which were disconnected from the battery. The electrons rush to equalise their potential, and the air atoms relax back into their normal state.

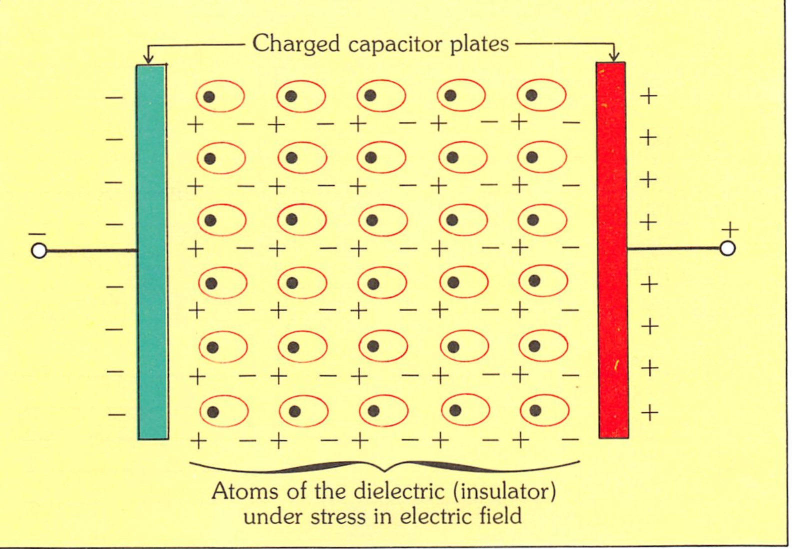
1. An electric field generated between two metal plates.

2. Energy stored in an electric field.

1



2



Measuring capacitance

The amount of electrical energy or charge which a capacitor stores depends on its physical dimensions, the gap between the plates, the nature of the insulating material between them (known as the **dielectric**), and the size of the potential difference applied to the plates. Below are some definitions and facts which are used to calculate capacitance.

Electric charge is measured in coulombs, as explained in a previous chapter, and the charge (Q) held in a capacitor is equal to the capacitance (C) of the capacitor multiplied by the voltage (V) applied across it:

$$Q = C \times V \text{ or } C = \frac{Q}{V}$$

Capacitance is measured in farads (F). One farad is the capacitance of a capacitor which stores 1 coulomb when a potential difference of 1 V is placed across it. Normally capacitance would be expressed in picofarads or microfarads, see *Table 2*.

The capacitance of a pair of parallel flat

plates is directly proportional to the area of the overlapping plates (S); inversely proportional to the distance between the plates (d); and directly proportional to the **permittivity** (symbol ε) of the dielectric (see *figure 3*). Thus:

$$C = \varepsilon \times \frac{S}{d}$$

The **relative permittivity** (formally called the dielectric constant) ε_r, is the number of times the permittivity of a material is greater than that for air, ε_o ($\varepsilon_o = 8.86 \times 10^{-12}$ farads per metre $= 8.86\ \text{pFm}^{-1}$). This gives us a number that is easier to handle than the actual constant. Some relative permittivities are given in *Table 1*.

Capacitors can be manufactured to any required size simply by altering these various dimensions and properties. Normally the aim is to get the capacitance required into the smallest possible space and many different materials and methods of construction are used.

Here is a worked example using the above formulae. Imagine two metal plates, each of area 400 cm^2, placed a distance of 5 mm apart in air. A potential of 500 V is then connected to them. The capacitance C can be found by the formula:

$$C = \varepsilon_o \times \frac{S}{d}$$

$$= 8.86 \times 10^{-12} \times \frac{0.04}{0.005}$$

$$= 7.088 \times 10^{-11}\ \text{F or } 70.88\ \text{pF}$$

The charge can be found by the formula:

$$Q = C \times V$$

$$= 7.088 \times 10^{-11} \times 500$$

$$= 3.544 \times 10^{-8}\ \text{C}$$

Disconnecting the plates from the supply and moving them together, so that d = 2 mm, means that C and V are now different:

$$C_1 = \varepsilon_o \times \frac{S}{d_1}$$

$$= 8.86 \times 10^{-12} \times \frac{0.04}{0.02}$$

$$= 1.772 \times 10^{-11}\text{F or } 17.72\ \text{pF}$$

$$V_1 = \frac{Q}{C}$$

$$= \frac{3.544 \times 10^{-8}}{1.772 \times 10^{-11}}$$

$$= 2000\ \text{V}$$

Table 1

Dielectric constant or relative permittivity

Material	r	Material	r
Amber	3	Clear mica	6
Bakelite	6	Micanite	5
Brown paper	2	Transformer oil	2.5
Waxed paper	4	Paraffin wax	2
Celluloid	3	Porcelain	5
Ebonite	3	Plate glass	6
Vulcanized rubber	3	Shellac	3

Notice that V has risen and C has dropped. Suppose now that a piece of glass with a relative permittivity of 6 ($\varepsilon = 6$), 20 mm thick, is placed between the two plates. C and V will again be different:

$$C_2 = \varepsilon \times \frac{S}{d}$$

$$= 6 \times 1.772 \times 10^{-11}$$

$$= 1.0632 \times 10^{-10}\ \text{F}$$

The charge between the plates is now:

$$V_2 = \frac{Q}{C_2}$$

$$= \frac{3.544 \times 10^{-8}}{1.0632 \times 10^{-10}}$$

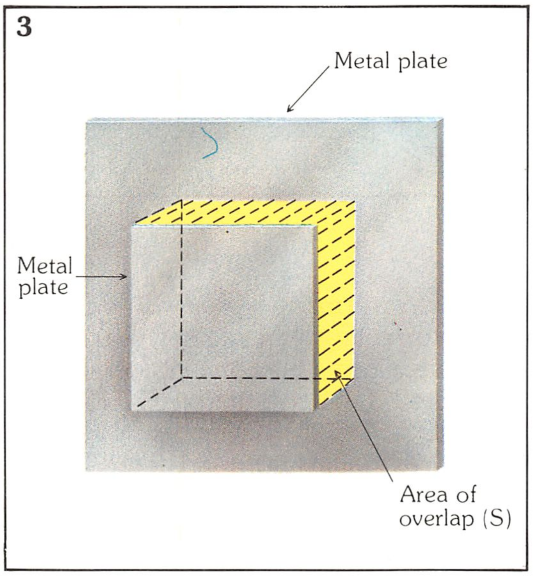
$$= 333\ \text{V}$$

Table 2

Multiples and submultiples of 10

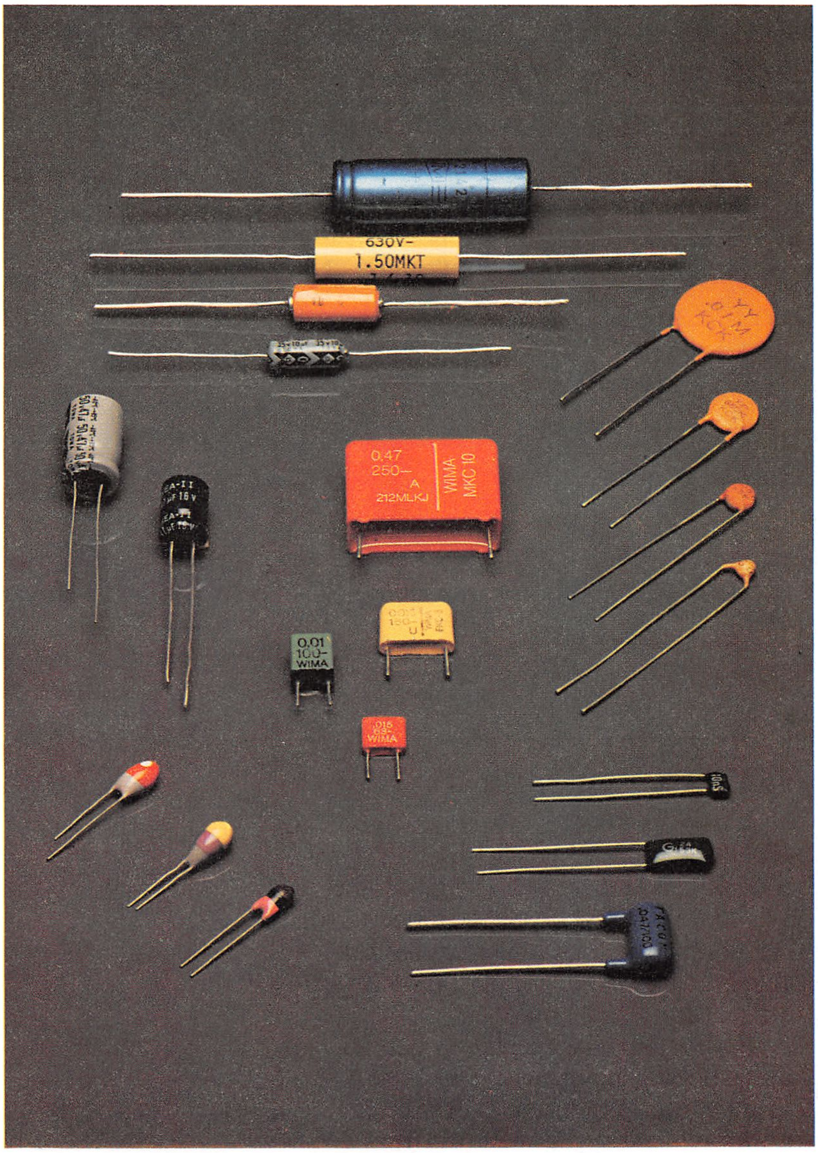
Abbrev.	Prefix		
T	tera	10^{12}	1,000,000,000,000
G	giga	10^{9}	1,000,000,000
M	mega	10^{6}	1,000,000
k	kilo	10^{3}	1,000
h	hecto	10^{2}	100
Da	deca	10	10
		1	1
d	deci	10^{-1}	0.1
c	centi	10^{-2}	0.01
m	milli	10^{-3}	0.001
μ	micro	10^{-6}	0.000 001
n	nano	10^{-9}	0.000 000 001
p	pico	10^{-12}	0.000 000 000 001

MΩ	= megohm	= 1,000,000 Ω
kV	= kilovolt	= 1,000 V
mA	= milliampere	= 0.001 A
μF	= microfarad	= 0.000 001 F
nF	= nanofarad	= 0.000 000 001 F
pF	= picofarad	= 0.000 000 000 001 F



3. The area of overlap between these two plates is where an electric field would operate.

Various types of capacitors: electrolytic, ceramic, plastic, tantalum and mica.



This time you will notice that C has increased and V has dropped, due to the change in the nature of the dielectric.

Charge and discharge

Because a capacitor has the ability to store energy it can, in practice, be used as a source of power for a limited length of time, after which it will have to be recharged. It has two phases, one of charge and one of discharge.

A capacitor does not actually allow electrical current to flow because the circuit is not continuously conductive; it contains an insulator, or **dielectric.** However, there is continuity in the phases of charge and discharge. If we apply an EMF to the plates, this will produce a potential difference across them and thus an electric field. This corresponds to the formation of an equal and opposing charge on the plates themselves. Thus by introducing a source of EMF into the circuit, electrons discharge from one plate to another, creating current.

If we now connect the two plates to a circuit, the electrons will tend to redistribute themselves again. Current will again flow, but in the opposite direction to the previous current. This will last until the electrons are uniformly redistributed across the metallic plates. The accumulated charge thus disappears, as does the voltage between the plates and the electric field. The capacitor is again discharged.

To analyse the movement of the current and the voltage flow of this change, the problem becomes more complicated. We could say that the charged current is high at the moment when the EMF is connected; this will only be limited by the resistance of the external conductors and the capacitor. When the voltage on the plates of the capacitor increases, the current is progressively discharged, then capacitor will block the current.

While the capacitor is discharged, the voltage is reduced in proportion to the discharging current. The capacitance and voltage will simultaneously reach the zero value.

The time taken for a capacitor to be charged or discharged will depend on its capacitance (C) and the resistance (R) in the charging circuit. By varying the resistance, this time can be increased or decreased.

The time taken to charge or discharge the capacitor is measured as the **time constant** (γ), which indicates how long it takes for the capacitor to be charged or discharged to a value corresponding to 63% of the applied voltage. After a period of time roughly equal to $5 \times \gamma$, any capacitor is completely charged or discharged. The time constant is calculated by:

$$\gamma = R \times C$$

DIGITAL ELECTRONICS

CMOS circuits

CMOS specifications

Bipolar transistor (both n-p-n and p-n-p) integrated logic circuits were first sold at the beginning of the sixties and although MOS transistors had been developed by this time, various technical difficulties prevented their immediate manufacture. These obstacles were overcome by the mid-1960s when the P-MOS and N-MOS logic families were introduced, to be followed in 1968 by the first series of CMOS logic families from RCA.

CMOS devices, using both p and n-type MOSFETS, are highly popular due to their wide tolerance of input voltages, good noise immunity and very low power dissipation. The latter property allowing engineers to put a great number of gates into a single IC, with consequent reduction in size and cost.

There are many types of CMOS device currently on the market, the most widely used being the conventional (or unbuffered) CMOS, buffered CMOS and the CMOS/SOS. Conventional CMOS devices are available in two series: the 74C and the 4000A. The 74C series is arranged with the same pin configuration as the TTL 74 series, while the 4000A is the original RCA series of 1968.

The 4000A series, also known as 14000A or 34000A, has the highest number of SSI and MSI circuits and operates on a recommended supply voltage of 3 to 12 V. Within the operating temperature range (−40° to 80°C, for plastic packages) the input current is 1 mA.

The layout of the buffered CMOS series is shown in *figure 1*. This series usually has the suffix B (e.g. 4000B). Its recommended supply voltage is from 3 to

Table 1
Static electrical characteristics of CMOS 'A', 'B' and 74C series

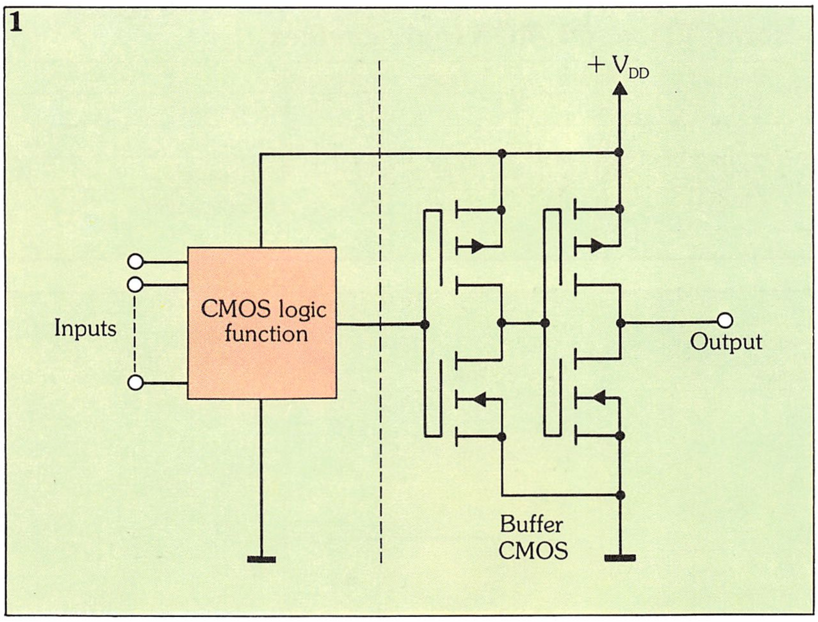
Symbol	Parameter	Conditions				Limits			Units
		V_{in}	Min	Max	V_{DD}	Min	Typ	Max	
V_{OL}	Voltage (Output Low)	5	—	—	5 (74)		0	0.05	V
		10	—	—	10 (A)		0	0.05	V
		15	—	—	15 (B)		0	0.05	V
V_{OH}	Voltage (Output High)	0	—	—	5 (74)	4.95	5	—	V
		0	—	—	10 (A)	9.95	10	—	V
		0	—	—	15 (B)	14.95	15	—	V
V_{NML}	Voltage (Noise Margin Input Low)	—	4.5	—	5 (74)	1	—	—	V
		—	9.0	—	10 (A)	1	—	—	V
		—	13.5	—	15 (B)	1	—	—	V
V_{NMH}	Voltage (Noise Margin Input High)	—	—	0.5	5 (74)	1	—	—	V
		—	—	1.0	10 (A)	1	—	—	V
		—	—	1.5	15 (B)	1	—	—	V
P_D	Power (Dissipation-DIP Plastic Package/25°C)	—	—	—	5 (74)	—	—	500	mW
		—	—	—	10 (A)	—	—	500	mW
		—	—	—	15 (B)	—	—	500	mW
P_T	Power (Dissipation per Output Transistor)	—	—	—	5 (74)	—	—	100	mW
		—	—	—	10 (A)	—	—	100	mW
		—	—	—	15 (B)	—	—	100	mW

Note: V_{DD} = 15 V is only valid for series B devices

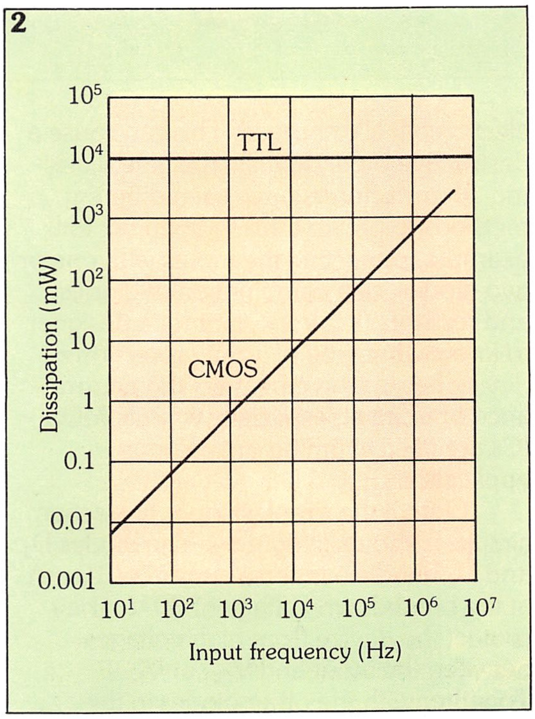
18 V and the maximum input current is 1 mA at 20 V, within the range of operating temperatures. *Table 1* summarizes the electrical and operating characteristics of the three CMOS families, CMOS A, B and 74.

CMOS/SOS (silicon on sapphire) gates are an example of the latest MOS technology. The substrate is made of sapphire and such gates have rise and fall times of less than 20 ns.

1



2



1. Layout of the buffered CMOS series.

2. Comparison of power dissipation versus frequency for typical CMOS and TTL gates.

CMOS power supplies

The main characteristic of a CMOS circuit is its low power consumption, however, the amount of power needed by the circuit will increase as the switching frequency rises. This is shown in *figure 2*.

As CMOS devices operate over a wide range of voltages and have a high noise immunity they do not need very sophisticated power supplies. However, it is important that the internal impedance of the supply is kept as low as possible to avoid interfering with the noise immunity of the circuits. This can be achieved by connecting a large capacitor in parallel with the supply.

CMOS circuits can tolerate a wider range of supply voltages than any other logic devices: TTL circuits run on 4.75 to 5.25 V, which is very narrow compared with the 3 to 18 V range of CMOS circuits. They are therefore particularly suitable for use in battery operated equipment, however, as supply voltage affects operating speed the lowest possible voltage supply compatible with the system's switching frequency must be used.

When switching, CMOS devices absorb a tiny transitory current pulse from the supply. This can cause a transient fall in the supply voltage as a result of the voltage drop across the resistance and inductance of the earth and supply lines. Although these devices have high noise immunity, this sort of disturbance can cause some problems if the leads are relatively long. To remedy this, connect a capacitor between the earth and supply terminals of each device.

CMOS operating temperatures

Another advantage that CMOS has over bipolar devices like TTL, is that of its operating temperature range. *Table 2* compares the operating temperatures of some TTL and CMOS NAND logic gates. The plastic packaged CMOS gates work correctly within a temperature range which is 40°C below, and 10°C above, that of the equivalent TTL gates.

The fan-out and drive capacity of a CMOS output

Figure 3 shows the equivalent circuit for the output of a CMOS device. When the

output is high, the logic element acts as a current source. Because of the internal resistance R_{OH}, the output voltage is given by $V_{DD} = 10$ V, with an arbitrary loss of 0.5 V in output, which gives $V_0 = 9.5$ V. The size of the difference between V_{DD} and the output voltage depends on the type of load being driven. The power dissipated internally by the device is equal to this difference multiplied by I_L.

R_{OH}, from 100 to 1000 Ω varies with the value of supply voltage, temperature, etc. It is different in each individual device and represents the resistance of one or more p-channel MOSFETs in series or parallel, turned on.

Figure 3b shows that when V_0 is low, the output absorbs (sinks) current through R_{OL}. The range of R_{OL} is similar to that of R_{OH}. If R_{OL} is taken to be 500 Ω, the value of the other circuit elements can be calculated:

$I_L = 1$ mA and $R_L = 9.5$ kΩ

These are fairly typical values.

When the device driven by the output is another CMOS gate, the fan-out is extremely high – typically 50. This is because the input of each CMOS device can be regarded as a capacitive load of a few picofarads. If CMOS outputs are used to drive devices and circuits other than CMOS, the fan-out may be limited. Manufacturers advise against the direct loading of CMOS devices with lamps or relays.

Short circuit in output

CMOS devices have an internal current limiting capacity. This safeguards the integrated circuit from any temporary short circuits between the output terminal and earth or V_{DD}. A continuous short circuit, however, can damage the device through overheating – especially when the supply voltage is high.

Input protection

The inputs of a CMOS device must be handled with care, as they can become electrostatically charged. This can cause a destructive breakdown in the gate dielectric. Manufacturers have used different methods to prevent this happening: for example, protecting the inputs with one or two diodes, or a combination of diodes and resistors. It is important for a designer to know what type of input protection a device has, as this can affect the performance of a circuit, especially when CMOS ICs are used in timing or analogue applications.

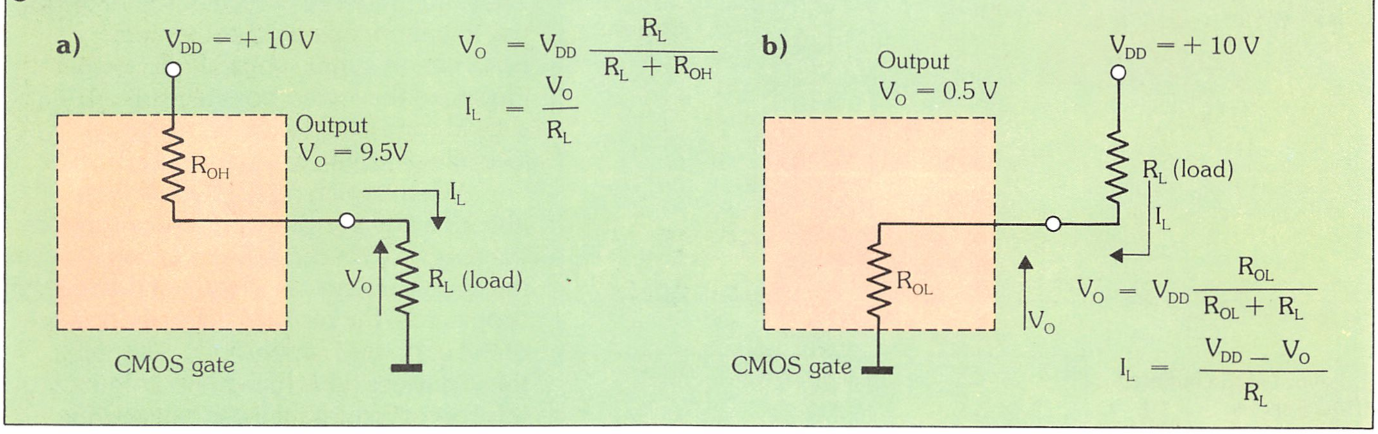
One of the simplest input protection circuits is shown in *figure 4*. The diodes D_1 and D_2 handle a reverse current of 10 mA at the breakdown voltage of 30 V. They protect the device from high voltages between the input and V_{DD} or V_{SS}. Together with the other diodes in the

Table 2

Comparison of operating temperature ranges for some TTL and CMOS logic devices

Family	Device	Temperature range
TTL	7400	0 to 70°C (industrial version)
TTL	5400	−55 to 125°C (military version)
CMOS	4011A	−40 to 80°C (plastic package)
CMOS	4011AD	−55 to 125°C (ceramic package)

3



3. The equivalent circuit for the output of a CMOS device, with output high (a) and output low (b).

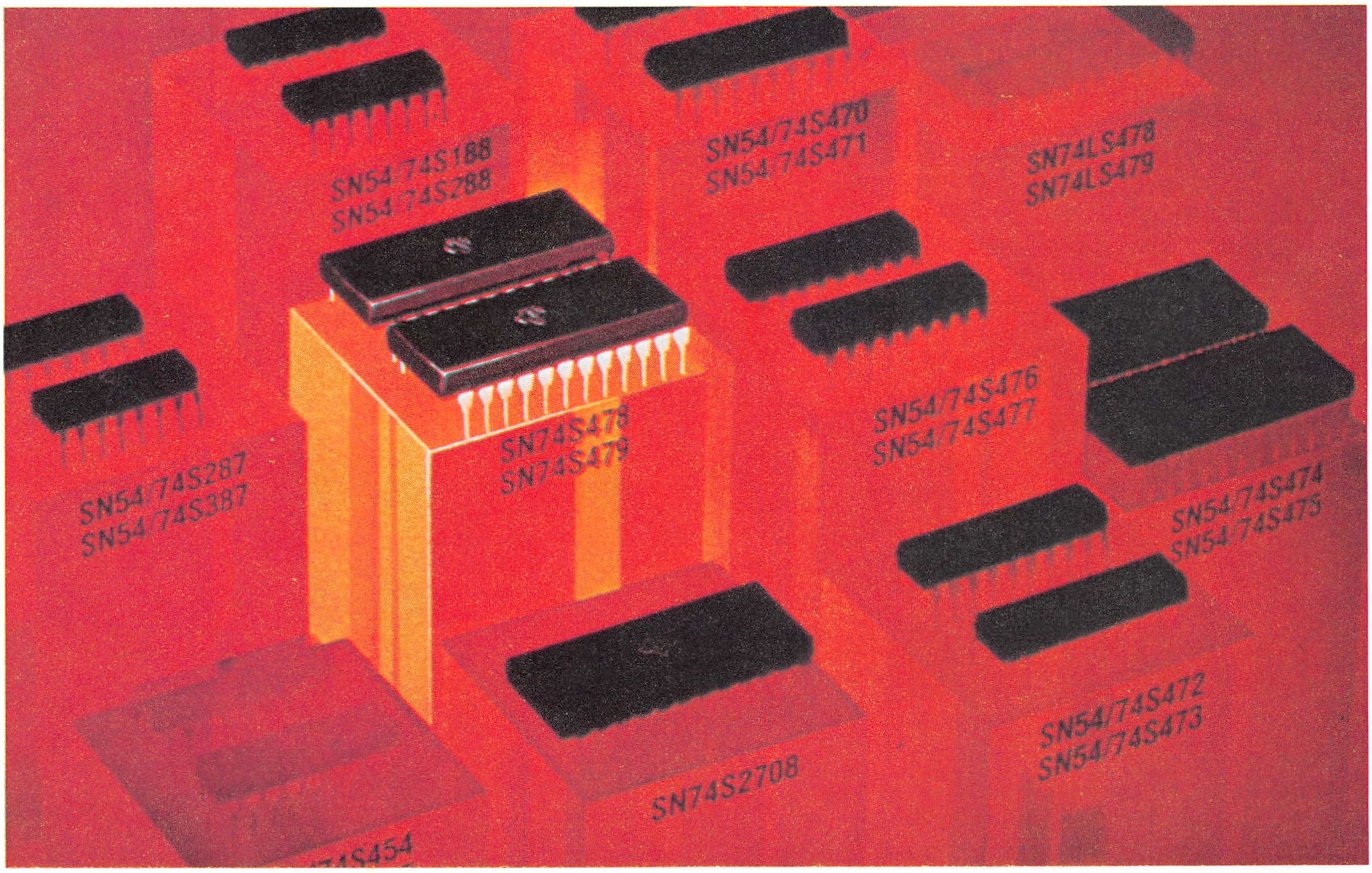
4. A simple input protection circuit using two diodes.

CMOS structure, they protect against high voltages between input and output.

Figure 5, shows another type of circuit which uses four protection diodes. Resistor R_1 limits the current, while diodes D_2 and D_3 work in breakdown mode. Resistor R_1 and diode D_1 do not exist as actual components within the IC, but are part of an MOS transistor arranged to act as a resistor and a diode.

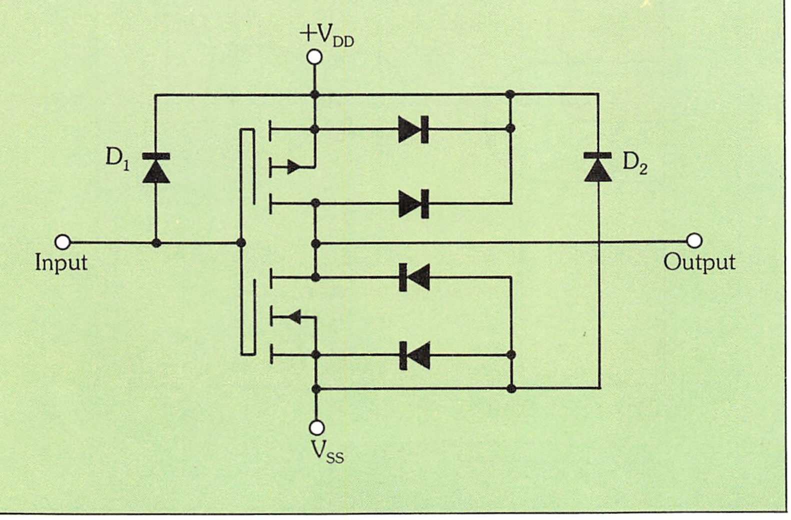
Unused terminals

CMOS gate inputs have a very high input impedance. Because of this, it is important that any unused inputs are connected as required to earth or V_{DD}. This prevents the accumulation of electrostatic charges which could affect the accuracy and working of the circuit. There is also a chance that without protection both the n and p-type MOSFETs could turn on simultaneously, thus overheating the device.



Bipolar ROM memories. These chips use TTL technology, but equivalent functions are available in CMOS.

4



Precautions to be taken when using CMOS devices

As we know, only a small amount of static electricity is enough to damage the insulation of a CMOS gate. Manufacturers recommend certain precautions to be taken to protect CMOS ICs from stray voltages:

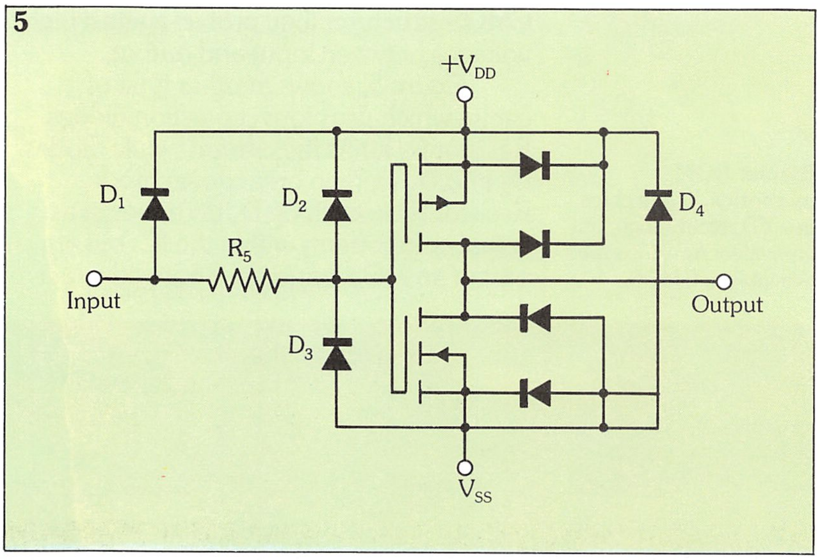
1) Make sure that the point of the soldering iron used in assembly is earthed;
2) Do not store the devices in plastic containers;
3) Disconnect test instruments before

switching off the supply;
4) If the input voltage could become negative, or greater than V_{DD}, then a protection resistor of at least 1 kΩ should be included in the circuit.

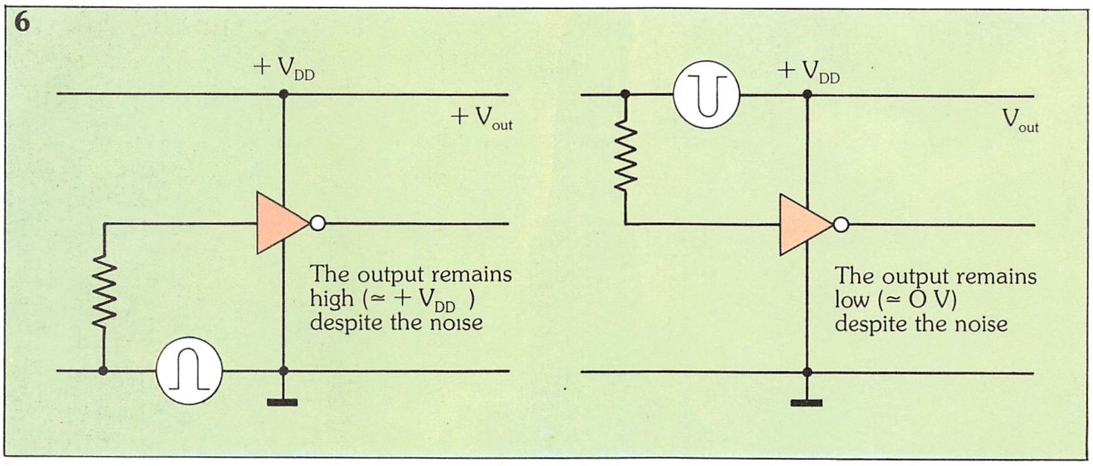
Immunity to disturbance

As we have seen, CMOS gates provide almost ideal input/output characteristics. Switching takes place when the input potential is about half V_{DD}. Typically the input signal can vary by 30% of the supply voltage without any significant variations in output (*figure 6*).

When high switching speed is not important, low speed functioning improves the noise immunity of a CMOS device. This is because slow circuits are less sensitive to the transitory electrical disturbances which are caused by energy stored in the small inductances and parasitic (unwanted) capacitances of the connecting leads.

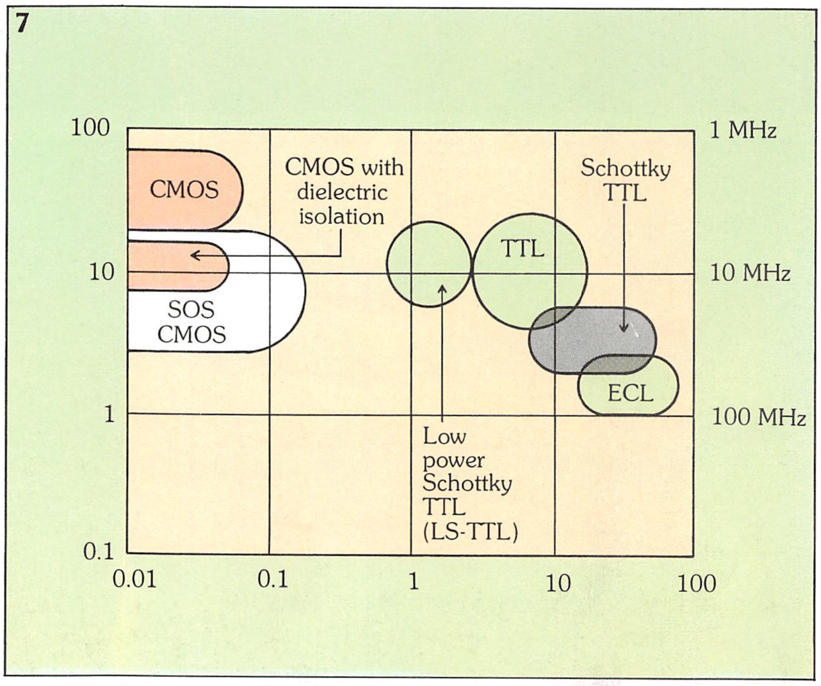


5. Input protection circuit using four diodes.



6. Noise immunity of a CMOS inverter.

7. The relationship between speed and power dissipation for gates of some of the most common logic families.



When to use CMOS logic circuits

CMOS ICs are superior to others in terms of immunity to disturbance, tolerance of supply voltage and stability of temperature and power dissipation: properties making them close to the ideal logic family. Limitations, however, such as low speed and low output current, mean that CMOS ICs cannot be used in all logic circuit applications.

Figure 7 shows the relationship between speed and power dissipation for gates of some of the most common logic families. Propagation delays and power dissipation are dependent on the operating

conditions as well as the gate design. Notwithstanding this, CMOS circuits should be used with a tolerated propagation time of between 20 and 90 ns; in other words the maximum switching speed should be less than 3 to 5 MHz.

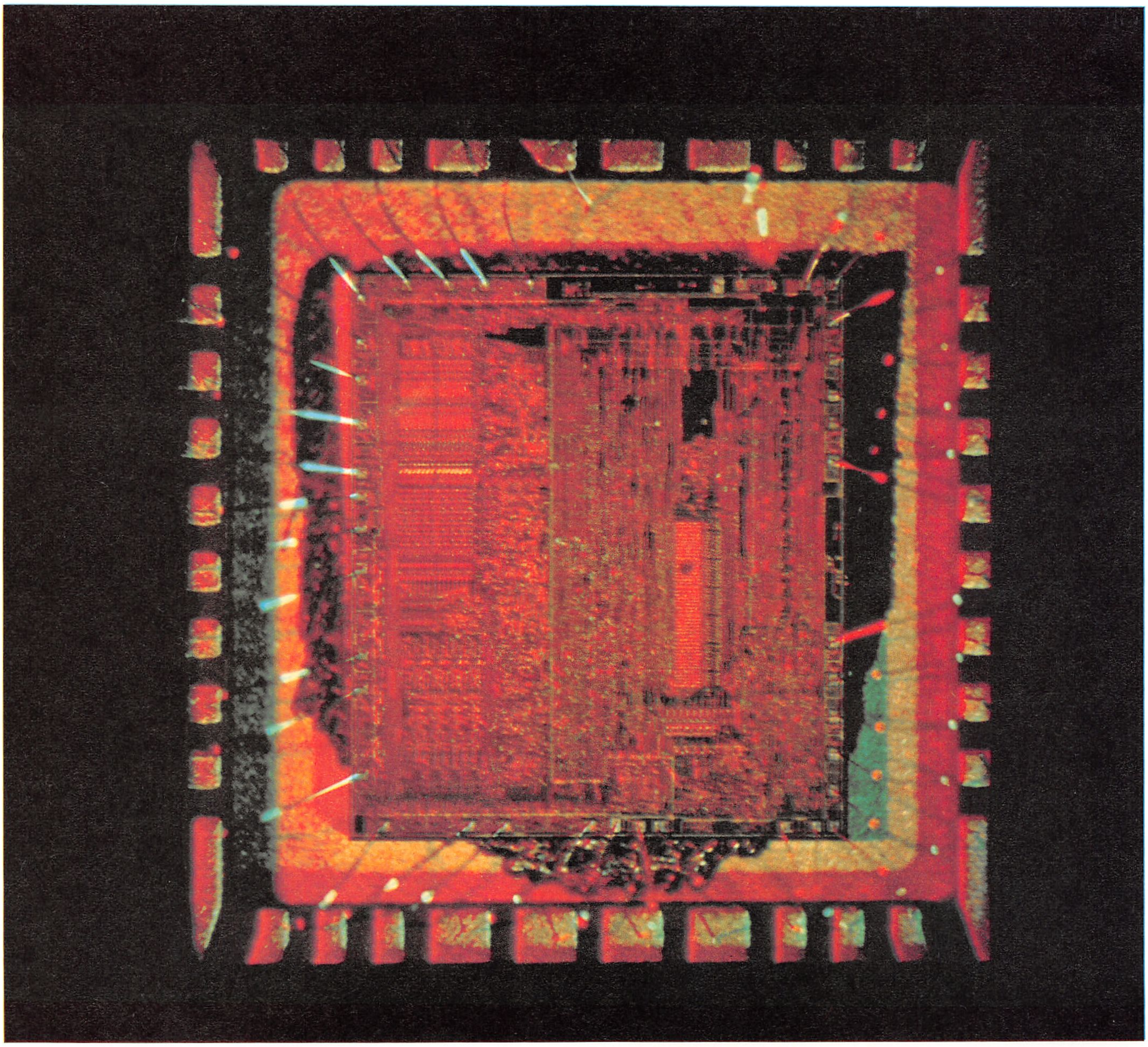
DI/CMOS (dielectrically isolated CMOS) and SOS/CMOS (silicon on sapphire CMOS) devices can be used in systems which require high speed switching. Unfortunately, these devices are not readily available on the market, have a limited choice of circuit functions and are very expensive.

In summary, we can see that CMOS integrated circuits are superior to other logic ICs, such as those of TTL and DTL families, in that they can operate from a wide range of power supply voltages. This, coupled with their low current consumption, makes tham an ideal choice for use in both battery powered and mains powered equipment.

Immunity to disturbance is high, and they will operate in a very wide range of ambient temperatures.

However, CMOS gates are inferior to those of other families in terms of switching speeds and output current, which means that they cannot be used in all applications.

The NCS 800 microprocessor which uses CMOS technology.



Interfacing between CMOS and TTL

Some CMOS manufacturers claim that their products are directly compatible with TTL, thereby making it possible to have the benefits of CMOS without redesigning a TTL circuit. One compatible characteristic of CMOS and TTL is that they will both operate within a voltage range of 4.75 V and 5.25 V. However, running a CMOS device on this sort of voltage level limits its speed of operation.

The CMOS 74C series of logic devices is directly compatible pin for pin with the TTL 74 series. This means that these TTL devices can be immediately replaced by CMOS gates.

But what would happen in a logic system when a TTL 74 is used in place of a CMOS 74C? In many cases the system will not operate correctly, even with speeds lower than those usually possible with CMOS. Replacing TTL 74 gates with CMOS 74C circuits, however, would probably function correctly and considerably reduce the power used.

The difference between these two results highlights the problem of interfacing between TTL and CMOS logic gates. In *figure 8*, the TTL 7400 gate represents the load for the equivalent CMOS 74C00 gate. As you know from the last chapter, the input of the 7400 requires a logic level 0 which is less than 0.5 V, and an input current of 1.6 mA (which flows from the load input to the output of the driving device). This means, that by Ohm's law, R_{OL} must be less than:

$$\frac{0.5\ \text{V}}{0.0016\ \text{A}} = 312\ \Omega$$

Unfortunately, the resistance of R_{OL} for a CMOS gate is not as low as this. So in the majority of cases, CMOS gates *cannot* drive TTL inputs correctly.

When CMOS devices *can* drive TTL gates it is at the expense of their noise immunity. Under these conditions, the CMOS output at logic 0 is close to the low level maximum input of TTL devices.

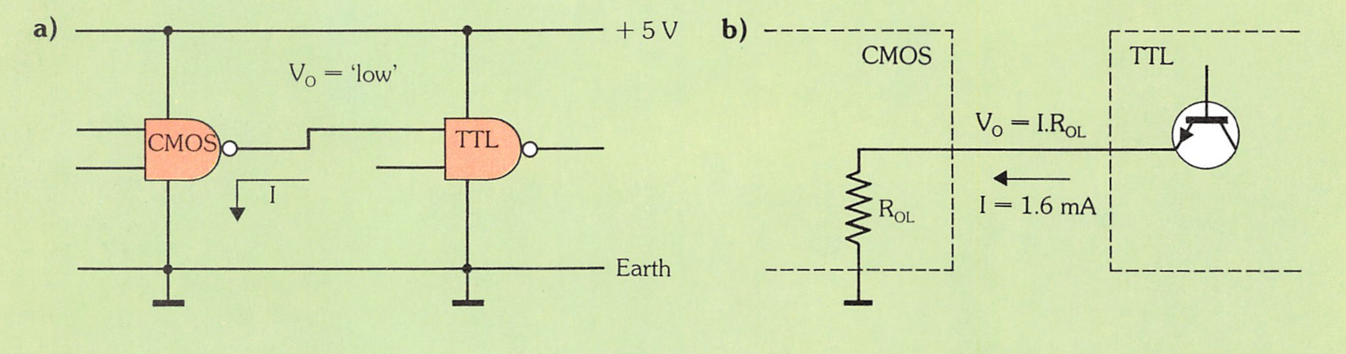
However, the inputs of TTL devices do not present any problems to CMOS gates at logic level 1. This problem is considerably simplified when the TTL gates belong to the L or LS series as they operate at lower input currents.

Figure 9 shows the current leaving the input of the 74L00 device to be only 0.18 mA at logic 0. This is considerably below the driving possibilities of the CMOS gate. From a practical point of view 74C devices can only drive *two* 74L inputs correctly; in other words they have a fan-out of two.

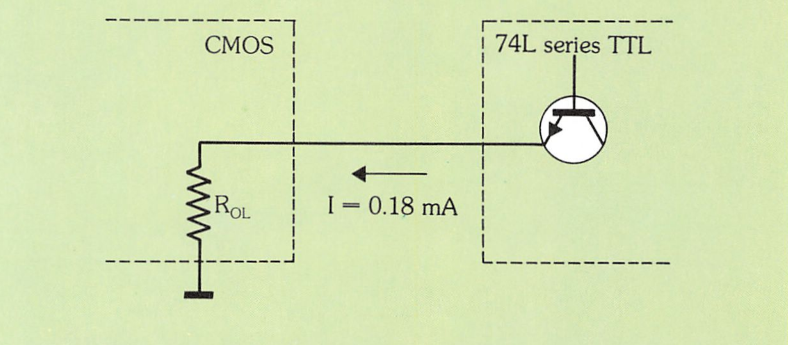
In some cases, which only concern

8. CMOS 74C devices are sometimes directly compatible with TTL circuits. Here a CMOS 74C00 drives a TTL 7400 IC (**b**) shows the equivalent electrical circuit.

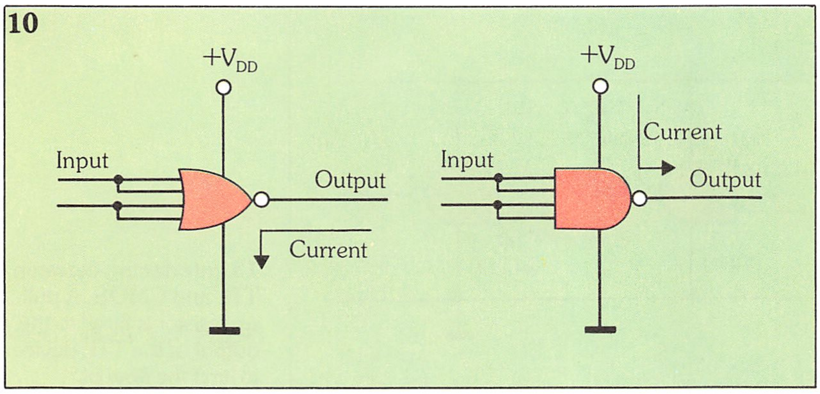
8

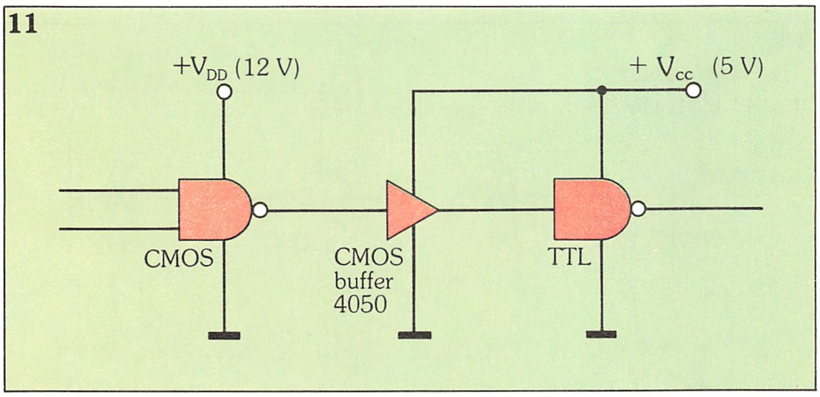


9



9. The same circuit as in 8 (b), but showing the current that flows – 0.18 mA which is below the driving capability of CMOS devices.

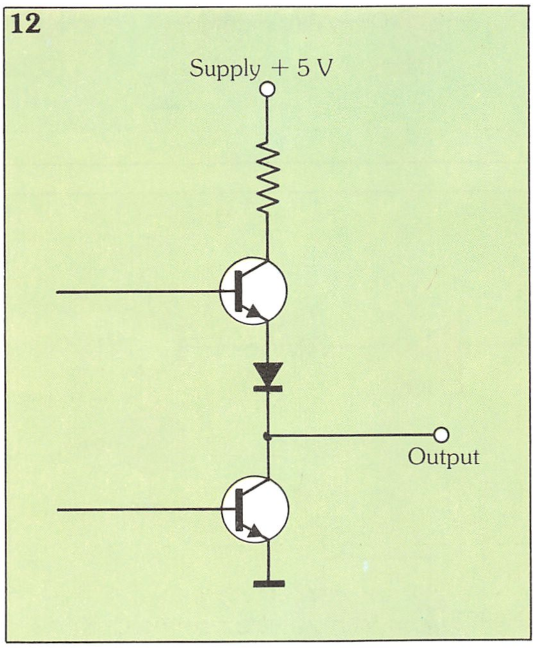




10. A 4-input CMOS gate used as a 2-input gate, to increase its output driving capability.

11. Interfacing CMOS and TTL with the 4050 buffer removes problems of compatibility between the two devices.

12. The totem pole output circuit of a TTL device.



CMOS devices in the A series, the output driving capacity at logic 0 can be improved by connecting some input terminals in parallel (*figure 10*). The increase in the NOR gate fan-out can be made clear if you look at the internal structure of the CMOS NOR gate shown in the previous chapter.

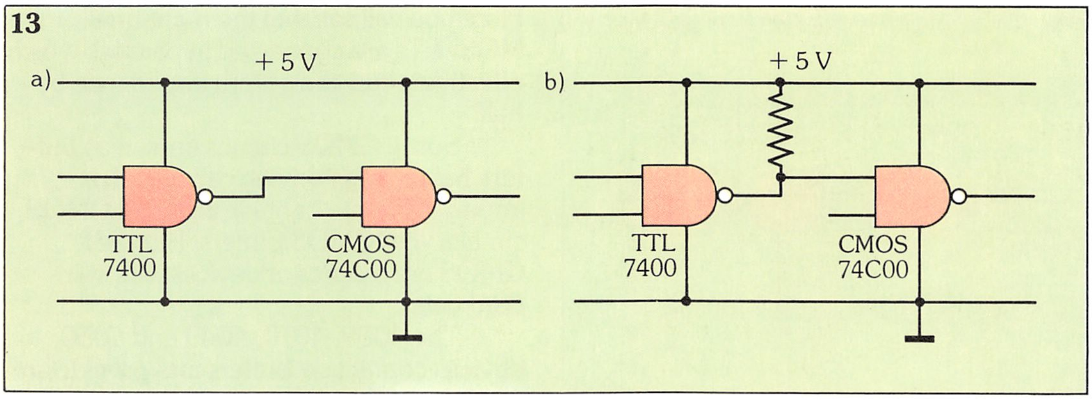
Here you will see that the n-channel MOSFETs are connected in parallel, which effectively cuts the output impedance by half.

Some CMOS circuits known as **buffers** have a high driving capacity. It is important not to confuse buffered CMOS circuits with CMOS buffers. Buffered CMOS are a series of devices with standard gates.

The 4009, 4010, 4049 and 4050 devices contain six buffers and provide an output current of 8 mA. This is large enough to drive five TTL inputs. Buffers can also be used for 'translating' or altering signal levels. Used as **interfaces** they allow CMOS circuits to operate with a higher supply voltage (from 6 to 15 V), while still allowing TTL circuits to run from their usual 5 V. *Figure 11* shows an example of a buffer gate from a 4050 CMOS IC being used to interface a CMOS gate to a TTL gate.

Finally, remember that the output of a buffered B series CMOS gate has greater driving capacity than an equivalent series A CMOS. Therefore, driving CMOS with a TTL presents no problems, as CMOS can operate correctly with 5 V supply and requires virtually no input current. However, sometimes a direct connection is insufficient. When a TTL output is high, it does not reach the level of the supply voltage. This is because part of the output signal is leaked through the transistor which is on, and the relevant series connected diode (*figure 12*).

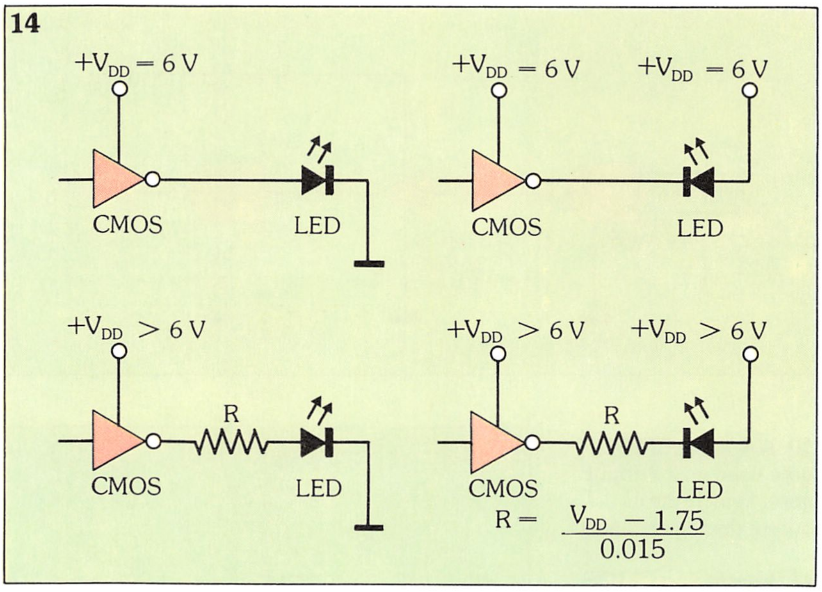
The output voltage can typically be less than 4 V. This is very close to the switching threshold of a CMOS input, if the device is run from a 5 V supply. Consequently, even if the CMOS device operates correctly, its noise immunity will suffer – any electrical disturbances could cause wrong switching to take place. To cure this, a pull-up resistor has to be fitted to the TTL output, as shown in *figure 13b*. Since the input of the CMOS device does not require current, the voltage drop in the external resistor is practically negligible and the TTL's logic 1 level will reach 5 V. This external pull-up resistor is necessary in a standard TTL as well as a low power TTL device because both types operate on the same voltage levels.

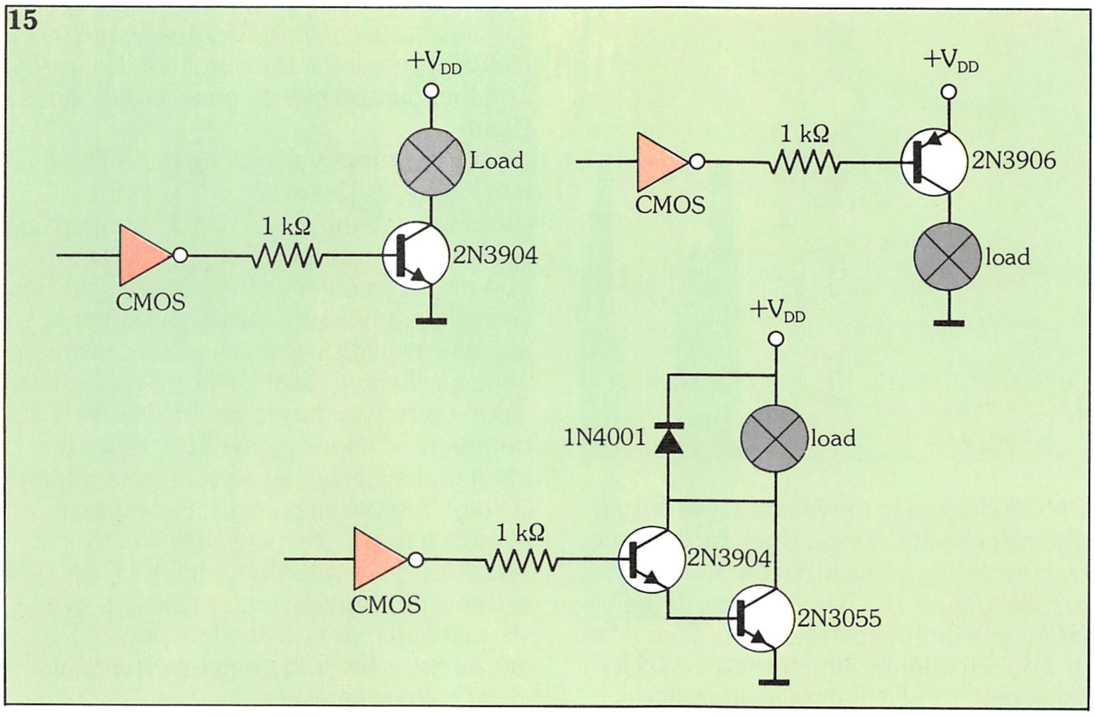


13. Interfacing between TTL and CMOS. A pull-up resistor is fitted to the output of the TTL device to limit the flow of current.

Driving capacity of a CMOS output

The CMOS output handles only a small amount of current, but this is enough to drive small loads. With a 6 V supply, the 4009, 4010, 4049 and 4050, along with the B series CMOS devices, can directly drive an LED. When higher supply voltages are used, a limiting resistor must be placed in series at the output (*figure 14*). When several gates drive an LED each, it is important to check that the power dissipated when they are all lit is not more than the CMOS device can handle. (Typically 200 mW.) Small incandescent light bulbs are an excessive load for a CMOS device. To drive this type of load, one or more additional transistors are needed, as shown in *figure 15*.





14. LEDs driven by CMOS.

15. CMOS outputs with additional driver transistors.

Some widely used CMOS ICs

4001 – 4025 – 4002 – 4078

The 4001 contains four, two-input NOR gates (*figure 16*). The 4025 contains three, three-input NOR gates. The 4002 contains two, four-input NOR gates and the 4078 has one, eight-input NOR gate.

These are used, either singly or in combination with other gates, in the design of complex logic functions. If needs be they can also be used as NOT gates by linking all the inputs in parallel.

The packages of these devices all have 14 pins. Pins 7 and 14 are the supply terminals V_{SS} (i.e. earth or 0 V) and V_{DD}. Any pins not connected are marked NC on the diagram.

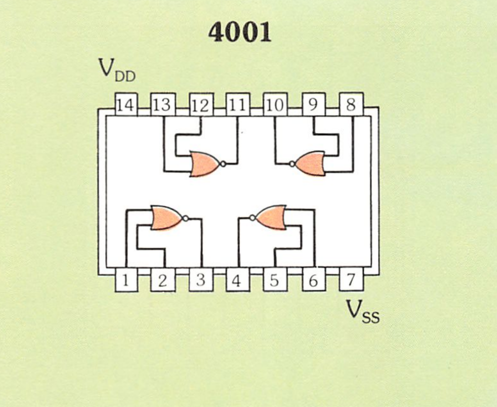
4071 – 4075 – 4072

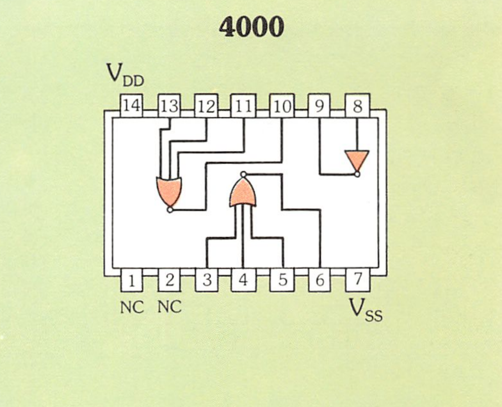
The 4071 contains four, two-input OR gates; the 4075 has three, three-input OR gates; and the 4072 contains two, four-input OR gates. *Figure 17* shows the schematic diagram for the 4071.

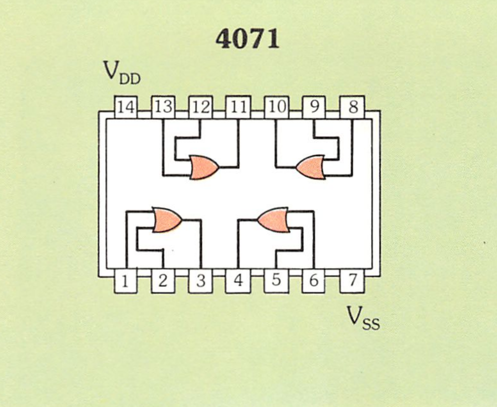
4000

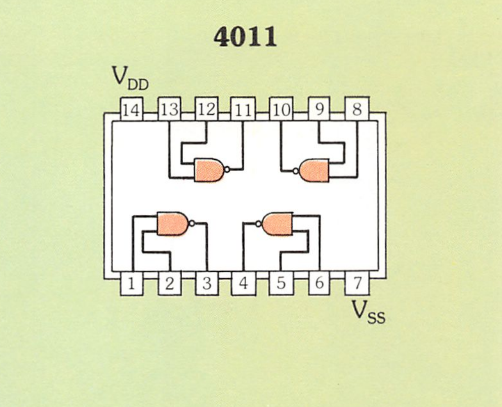
This IC contains two, three-input NOR gates and an inverter. These gates can be combined to make different logic functions: A NOR gate followed by an inverter becomes an OR gate. By applying this OR gate to one of the inputs of the other NOR gate, a five-input NOR can be made. This is shown in *figure 18*.

16, 17, 18, 19. Pin layouts of common CMOS ICs.

16


18


17


19


4011 – 4023 – 4012 – 4068

The 4011 has four, two-input NAND gates (shown in *figure 19*). The 4023 contains three, three-input NAND gates. The 4023 contains three, three-input NAND gates. The 4012 has two, four-input NAND gates and the 4068 is an eight-input NAND gate.

4081 – 4073 – 4082 – 4069

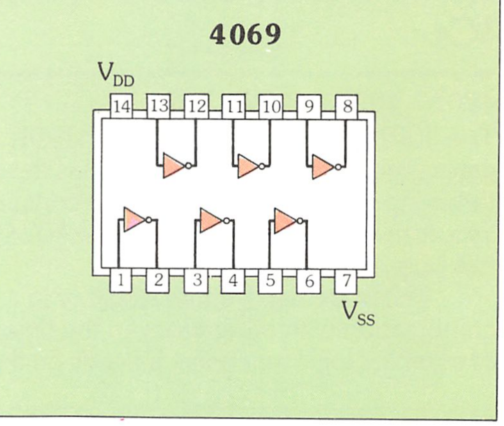
The layout of the 4081 is shown in *figure 20.* This contains four, two-input AND

gates. The 4073 IC has three, three-input AND gates and the 4082 contains two, four-input AND gates. The 4069 contains six inverters and is shown in *figure 21*.

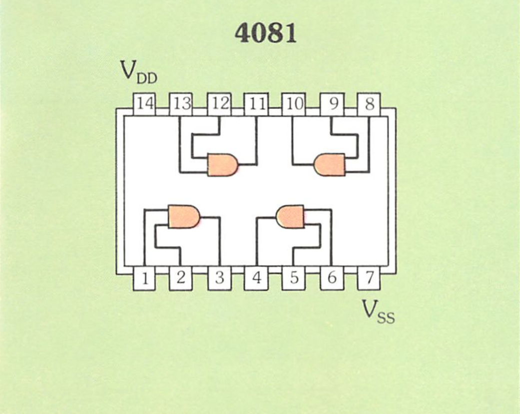
4049 – 4050

The 4049 can be used as a buffer or to interface CMOS to other logic families. The input voltage can be higher than the supply voltage. With a 5 V supply, each buffer can drive two TTL inputs. A single IC contains six buffers (remember, these invert their input signal). The 4050 has the same characteristics but does not invert the input signals.

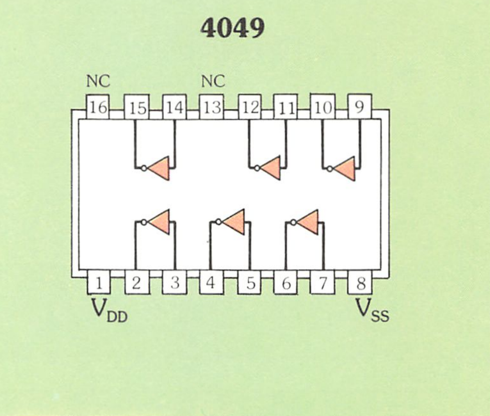
21



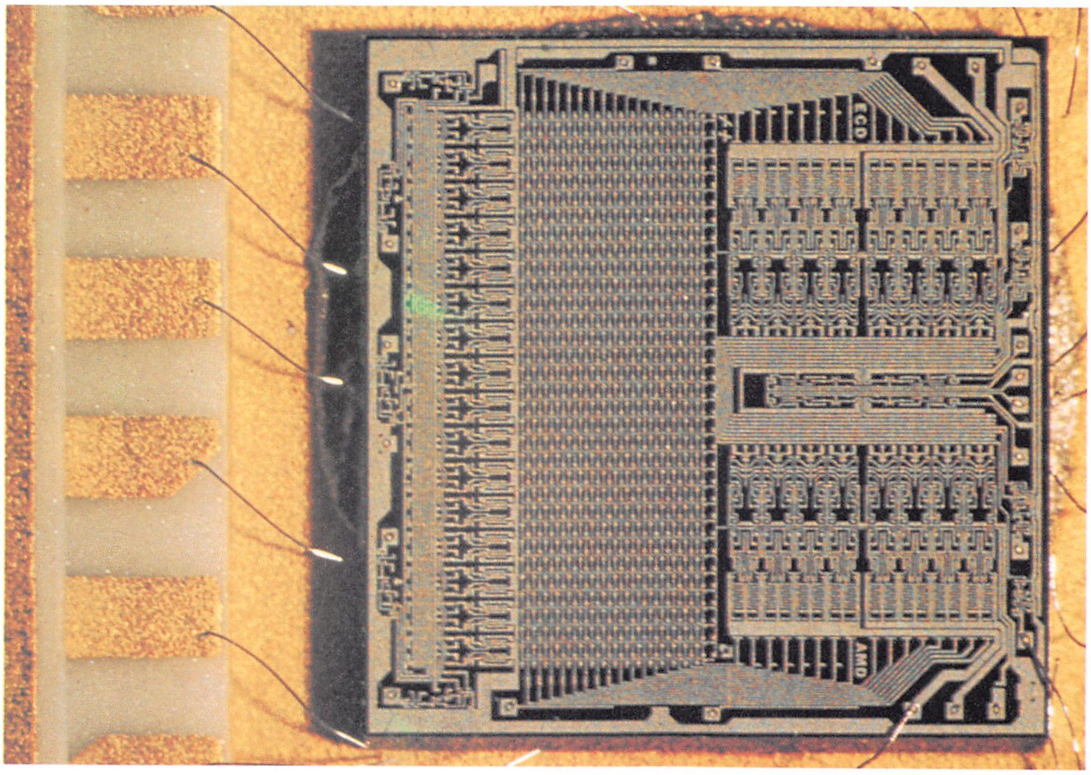
20



22

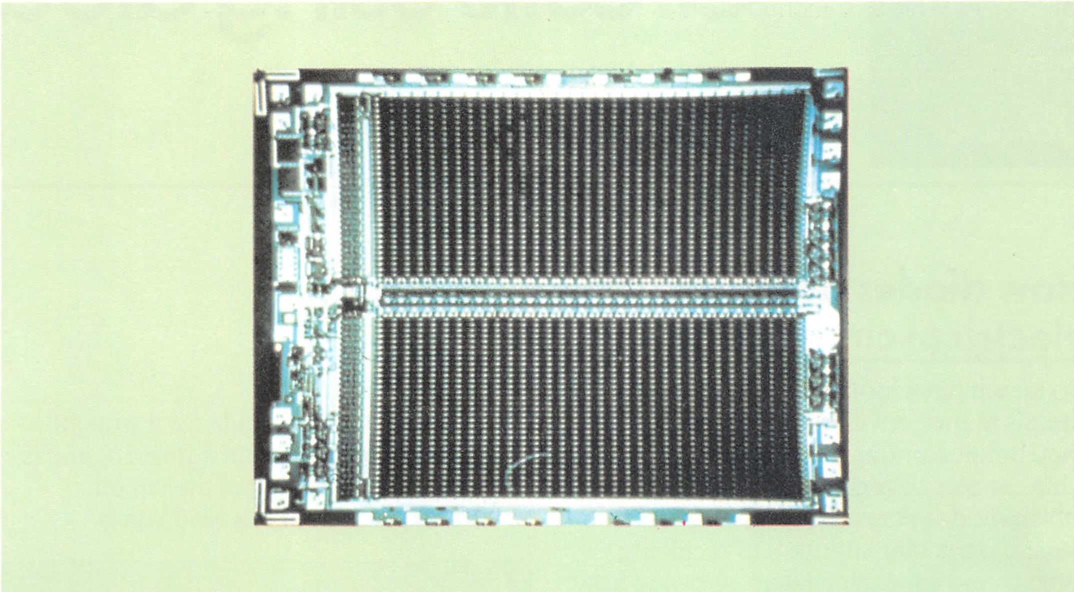


20, 21, 22. Pin layouts of common CMOS ICs.



Science Photo Library-James Bell

Close-up of an IC showing the connecting wires which run from the chip to the package terminals.



Photomicrograph of the TMS 2147 static MOS memory.

Glossary

buffered CMOS	CMOS logic circuits with larger output driving capacity than normal
CMOS	complementary MOS. Logic circuits using p and n-type MOSFETs working in complementary mode – when one is on the other is off
CMOS buffer	amplifying device used to boost outputs of CMOS logic devices
DI/CMOS	dielectrically isolated CMOS. Expensive, high speed device, with a degree of isolation between gates making it less prone to parasitic capacitances
MSI	medium scale integration. Device with between 10 to 100 gates on its chip
noise	unwanted stray voltages in a circuit which can cause a circuit to operate inaccurately
pull-up resistor	resistor added to a circuit to increase its output voltage
SOS/CMOS	silicon on sapphire CMOS. Very fast but expensive series of CMOS gates
SSI	small scale integration. Logic device with less than 10 gates on its chip

SOLID STATE ELECTRONICS

Circuits using diodes

How diodes behave in electrical circuits

So far we have looked at the many uses of diodes in modern electronics, and how they behave under an applied voltage. This chapter covers the main circuits in which diodes play an important part, paying particular attention to rectifying circuits and capacitive filters, showing how they are employed to solve the problem of using a.c. supplies in electronic circuits.

First, it's important to understand how the functioning of a diode can be related to other electrical components. The problem is that the resistance of a diode is variable, so an increase in voltage does not produce a proportional – or linear – increase in current. This non-linear relationship can be seen by the curve of a diode characteristic, part of which is shown in *figure 1b*.

From the shape of the curve we can see what the resistance is doing. With a linear device the resistance is constant – voltage and current increase in proportion to each other – and the line is straight. Here, given a voltage, a known constant resistance and the formula $V = I \times R$, the current can easily be calculated. Not so with the diode, however. *Figure 1a* represents a simple circuit comprising a diode and a resistor. It has an input voltage of V_i. We want to find the relationship between voltage and current for this circuit.

If the input voltage (V_i) is a known quantity, we can usually calculate voltage drop across the load resistor by using Ohm's law: $V_L = I \times R_L$. The real problem though is to calculate the value of the current, I.

The input voltage V_i is equal to the voltage drop across the diode V_d plus the voltage drop across the resistor:

$$V_i = V_d + I \times R_L$$

transposing we have:

$$I = \frac{V_i}{R_L} - \frac{V_d}{R_L}$$

This is, in fact, the formula for a straight line on a graph, shown in *figure 1b,* and is known as the **load line** of the circuit.

The position of this load line is dependent on the value of the input voltage.

Note that: the I axis intercept at point A occurs when:

$$I = \frac{V_i}{R_L}$$

the V axis intercept at point B occurs when:

$$V_i = V_d$$

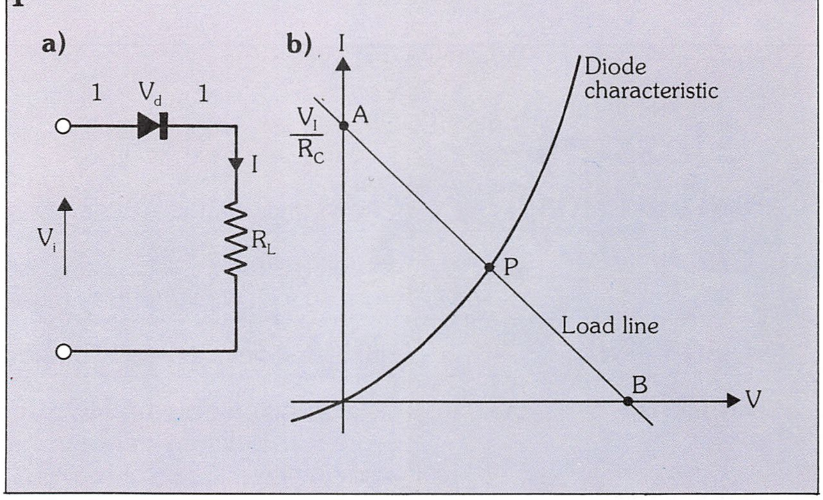
the slope of the straight line is:

$$-\frac{1}{R_L}$$

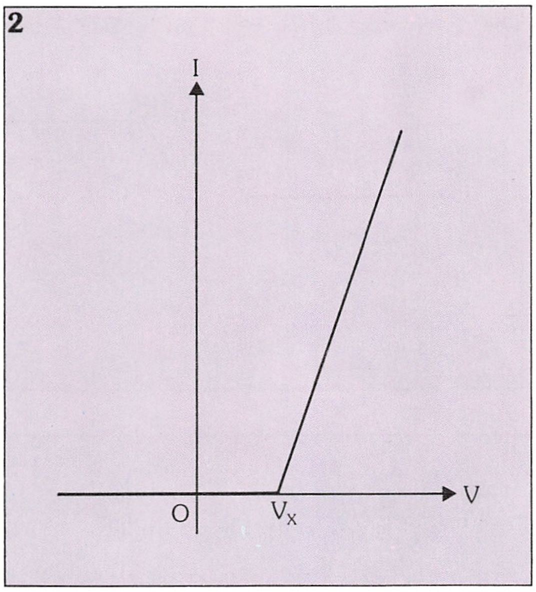
The point where the load line and the diode characteristic meet is the operating point P, whose co-ordinates give the current which will flow in the circuit and the voltage drop across the diode for a particular input voltage.

However, different input voltages

1



1. The circuit used to plot the electrical characteristics of diodes, and an example of the type of graph obtained.



2. A 'straight line' diode characteristic, used to approximate the behaviour of a device.

have different load lines and therefore different operating points.

This graphical method, showing us the relationship between voltage and current, makes it much easier to analyse the behaviour of electrical circuits containing diodes.

The straight line diode characteristic

When ascertaining how a diode will perform in a circuit, it is not always necessary to work with the exact I-V characteristic of the particular diode in question. A simplified, straight-line model, like the one in *figure 2,* can often be used instead, as an approximation of the actual characteristic.

Let's see how close this model comes to the actual characteristic. We know that the diode does not begin to function until it crosses the threshold voltage V_x, so for practical purposes the current can be considered to be zero until this voltage is reached. After this point, the model represents the diode as a conductor with a low, fixed resistance, as shown by the straight line. This gives a linear model to work from, where current is proportional to voltage. As we have said, this is only an approximation of what actually happens but it is close enough for us to work with.

When a reverse voltage is applied, the functioning of the diode is approximately that of an infinite resistance. The forward resistance is generally very low, and at times negligible.

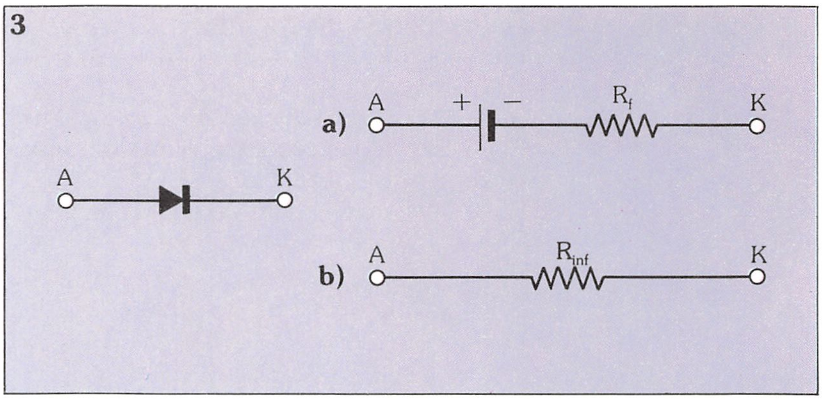
The use of this model makes the study of circuits with diodes much simpler. In fact there are only two operational possibilities. They either conduct (on) or they do not conduct (off). So if the voltage applied to a diode exceeds the threshold voltage, then the diode is on; when the voltage is lower than this, the diode is off.

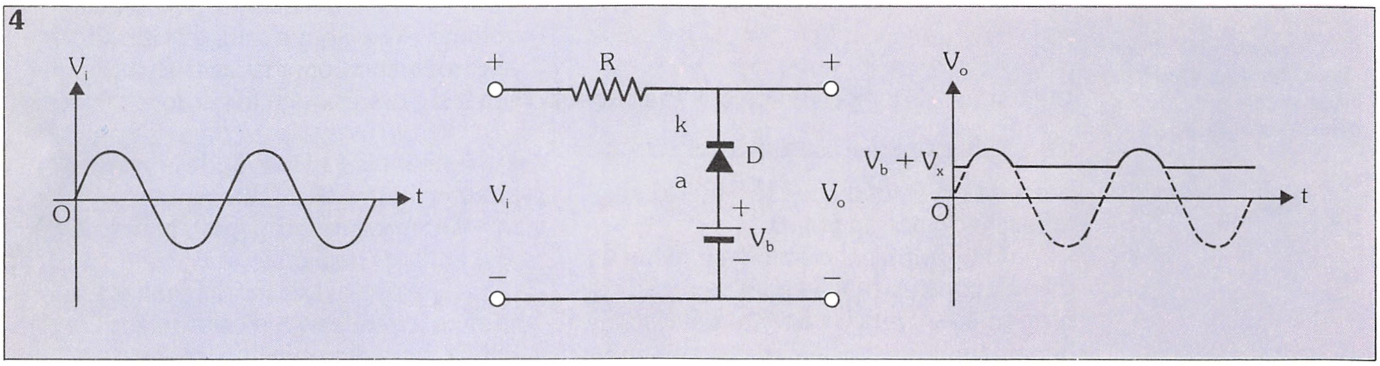
Figure 3 illustrates how the diode functions with forward and reverse vol-

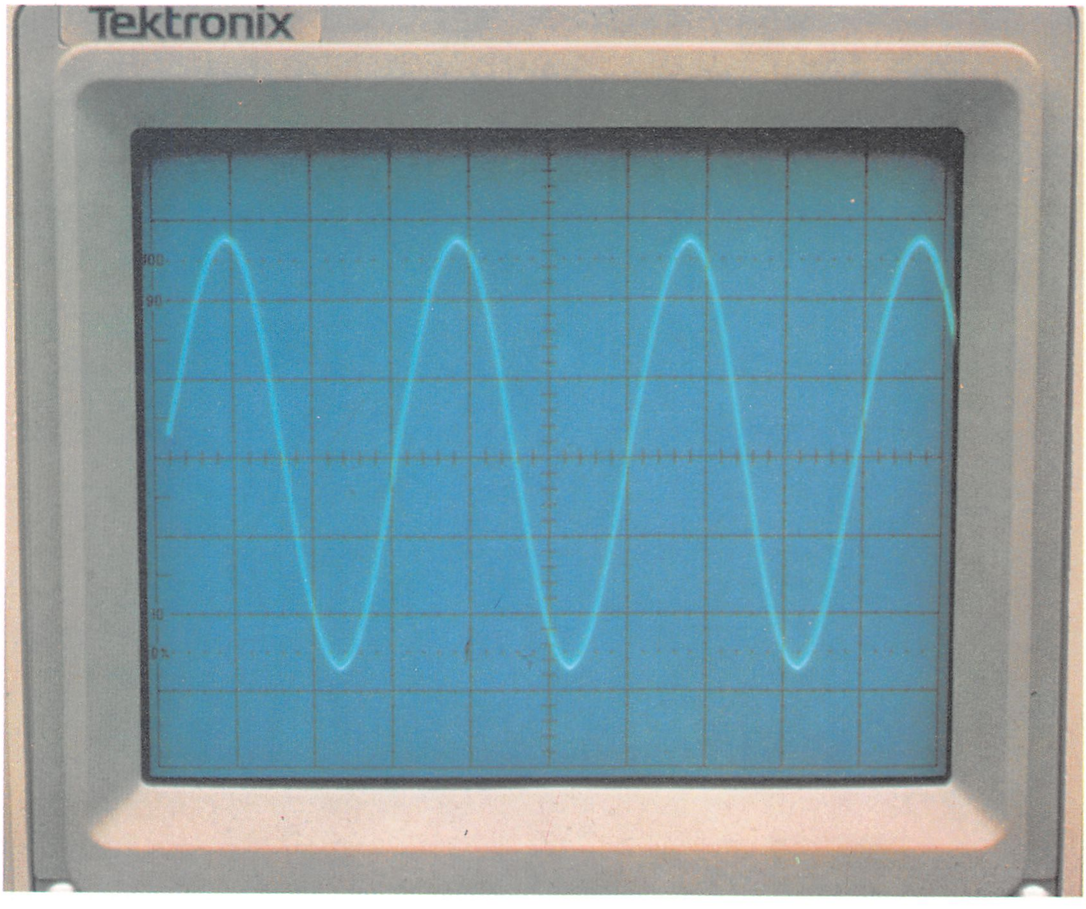
Sophisticated equipment like this oscilloscope is used to study electrical waveforms and the action of electronic devices. (Photo: Tektronix).

tages. That is, once past the threshold voltage we can consider the diode as a battery with a constant, low resistance (*figure 3a*); under reverse voltage we simply have an infinite resistance (*figure 3b*). This then, is a model of how the diode will work in the system.

Designing a system can be made more simple by thinking in terms of a model. Replacing the diode symbol in a circuit by its functional models (*figure 3a* and *b*) shows us, at a glance, how the diode performs at a given moment.







3. Models representing a conducting diode (a) and a non-conducting diode (b).

4. A limiting circuit, built with a diode. Only the part of the input waveform that is greater than $E_R - V_X$ is passed to the output of the circuit.

A sine waveform – like that of alternating current – displayed on an oscilloscope.

Limiting circuits

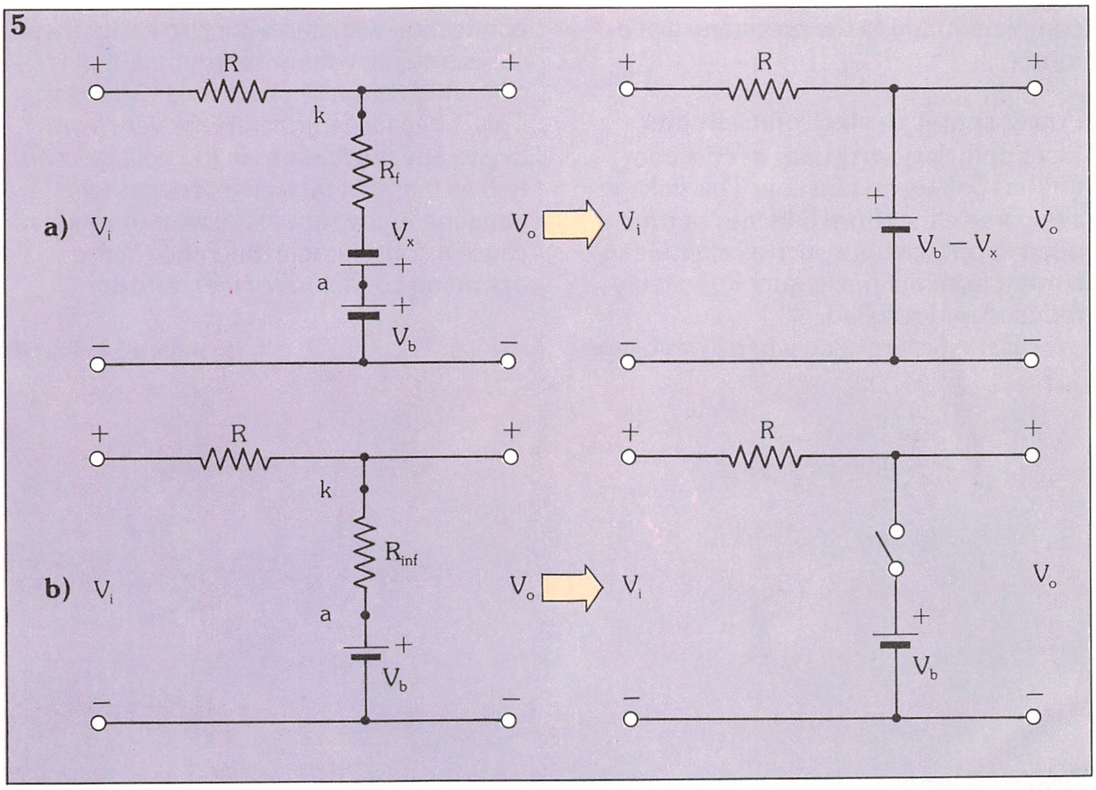
A limiting circuit, as the name suggests, limits the input wave forms and only allows signals which exceed a preset **reference level** to pass through the circuit.

An example of this is shown in *figure 4:* the circuit is composed of a resistor, a battery and a diode. The input voltage, V_i, comes from an alternating voltage source. The battery has a voltage V_b, which is set slighty higher than the desired reference level. This is to take into account the threshold voltage, V_x, of the diode.

The output voltage, V_o, is measured across the branch containing the diode and the battery. The positive pole of the battery is connected to the anode of the diode.

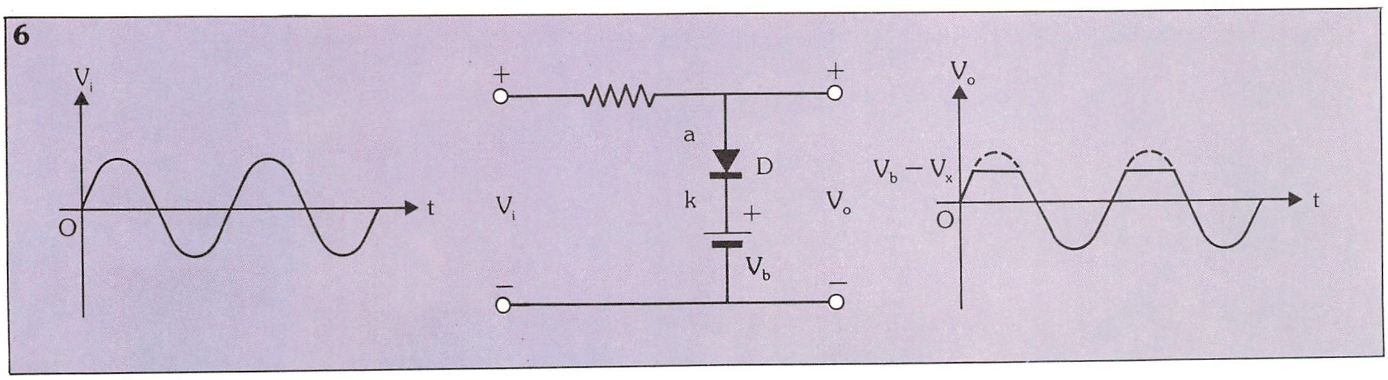
Replacing the diode by the equivalent functional model given in *figure 3a,* produces the circuit illustrated in *figure 5a,* when the diode is conducting. This can be simplified further: the two batteries can be incorporated to give a single voltage battery (V_b-V_x) and the resistance R_f can be ignored because it is an extremely low value. When the diode is in the off position, the resistance becomes infinite and can be replaced by an open switch (*figure 5b*).

Looking at the circuit you can see that as long as the input voltage, V_i, is less than the equivalent battery voltage (V_b-V_x), it cannot pass through the circuit, because



5. (a) Functional model for a diode limiting circuit when the diode is conducting. (b) shows the equivalent circuit for when the diode is not conducting.

6. Another diode limiting circuit. In this case voltages lower than $E_R - V_X$ are passed to the output.



the greater voltage of the battery will operate and the output voltage, V_o, will be the battery voltage. But when the input voltage, V_i, exceeds the battery voltage (i.e. $V_i > V_b - V_x$) the cathode of the diode will be more positive than the anode and goes to the off position; so the output voltage in this case will be the input voltage. The resulting waveform of V_o is shown in *figure 4*.

By simply changing the diode around and connecting its cathode to the positive pole of the battery, we will get a completely different waveform. Looking at *figure 6* you can see that with this arrangement the diode would conduct when it was previously switched off and vice versa. In this way we obtain a waveform which is complementary to the preceding waveform.

Power supply to electronic circuits

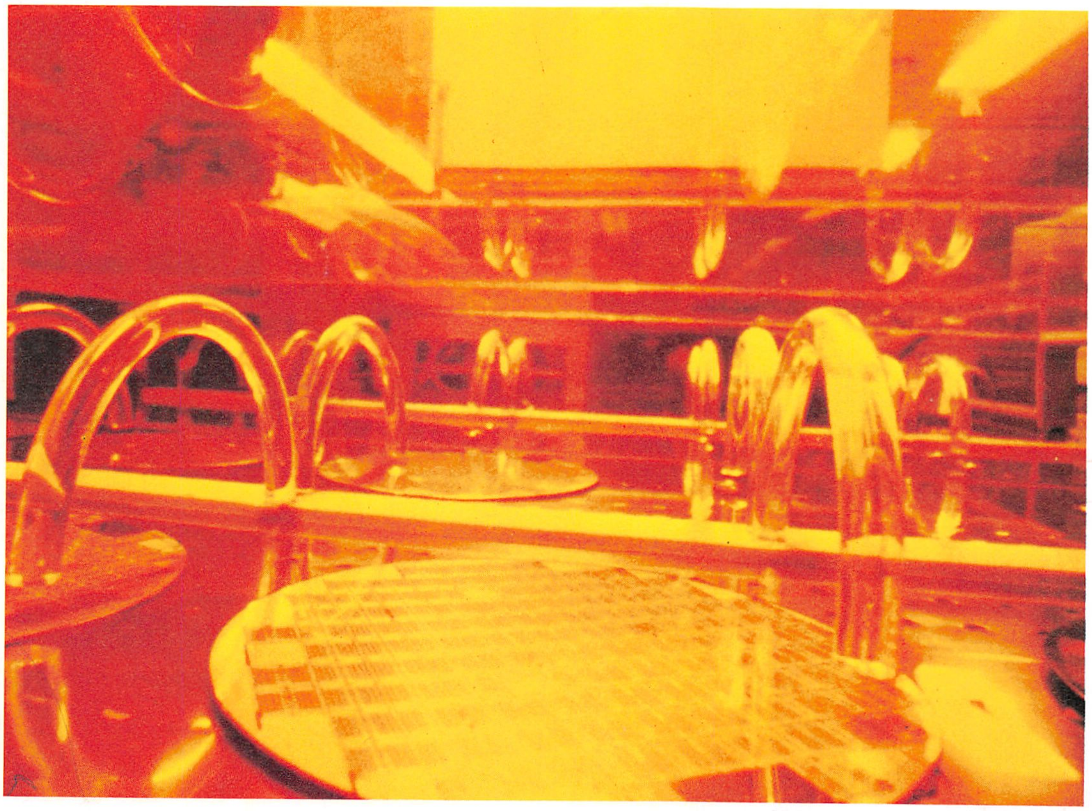
Electronic devices usually need a continuous voltage to function. This voltage can come either from batteries or from other supply circuits, such as alternating current from the mains supply, suitably reduced and rectified.

Batteries are used when low voltages (1.5 – 24 V) are required. Other supplies are used where high voltages are needed or in circuits which have a high current consumption. If a battery is used with a circuit that has high power consumption, it will have a very short life and so need to be replaced rapidly.

In general, batteries are used in portable equipment such as transistor radios, cassette players and so on. Other types of supply circuits are used in almost all other cases. There are various types of supplies. The most frequently used are those which convert the alternating voltage from the mains (220 – 240 V at 50 Hz) into the continuous voltage required for the circuit.

You will sometimes come across the term **stabilized supplies,** particularly in connection with transistor circuits. In these cases a stable voltage is required, that is, one which has no appreciable variations. This is because transistors are very sensitive to any fluctuations in the voltage used to bias them. In fact, even a very slight variation in the supply is usually enough to cause a considerable difference in the operating conditions of the transistor.

(continued in Part 8)



A stage in the manufacture of IC chips. Semiconductor slices are washed in a special bath. (Photo: courtesy IBM).